ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-882

Can States Sustain and Replicate School District Improvement? Evidence from Massachusetts

Beth E. Schueler
University of Virginia

Liz Nigro
University of Virginia

John Wang
University of Virginia

The improvement of low-performing school systems is one potential strategy for mitigating educational inequality. Some evidence suggests districtwide reform may be more effective than school-level change, but limited research examines district-level turnaround. There is also little scholarship examining the effects of turnaround reforms on outcomes beyond the first few years of implementation, on outcomes beyond test scores, or on the effectiveness of efforts to replicate district improvement successes beyond an initial reform context. We study these topics in Massachusetts, home to the Lawrence district representing a rare case of demonstrated improvements in the early years of state takeover and turnaround and where state leaders have since intervened in three other contexts as a result. We use statewide student-level administrative data (2006-07 to 2018-19) and event study methods to estimate medium-term reform impacts on test and non-test outcomes across four Massachusetts-based contexts: Lawrence, Holyoke, Springfield, and Southbridge. We find substantial district improvement was possible although sustaining the rate of gains was more complicated. Replicating gains in new contexts was also possible but not guaranteed.

VERSION: November 2023

Suggested citation: Schueler, Beth, Liz Nigro, and John Wang. (2023). Can States Sustain and Replicate School District Improvement? Evidence from Massachusetts. (EdWorkingPaper: 23-882). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/2psy-nw21

Can States Sustain and Replicate School District Improvement? Evidence from Massachusetts

Beth E. Schueler, Liz Nigro & John Wang
University of Virginia

Abstract: The improvement of low-performing school systems is one potential strategy for mitigating educational inequality. Some evidence suggests districtwide reform may be more effective than school-level change, but limited research examines district-level turnaround. There is also little scholarship examining the effects of turnaround reforms on outcomes beyond the first few years of implementation, on outcomes beyond test scores, or on the effectiveness of efforts to replicate district improvement successes beyond an initial reform context. We study these topics in Massachusetts, home to the Lawrence district representing a rare case of demonstrated improvements in the early years of state takeover and turnaround and where state leaders have since intervened in three other contexts as a result. We use statewide student-level administrative data (2006-07 to 2018-19) and event study methods to estimate medium-term reform impacts on test and non-test outcomes across four Massachusetts-based contexts: Lawrence, Holyoke, Springfield, and Southbridge. We find substantial district improvement was possible although sustaining the rate of gains was more complicated. Replicating gains in new contexts was also possible but not guaranteed.

Author Note: This research has been supported by the Spencer Foundation and WT Grant Foundation. Liz Nigro and John Wang's work on this project was supported by the Institute of Education Science, U.S. Department of Education through grant R305B200005 to the University of Virginia (UVA). The authors thank Ziqiao Chen for significant research assistance and data management, our colleagues at the EdPolicyWorks Center at UVA, Association for Education Finance and Policy 2023 conference participants, NAEd/Spencer 2023 Annual Fall Retreat participants, Doug Harris, Susan Moore Johnson, Kris Gutiérrez, Whitney Kozakowski, Jim Soland, and Kylie Anglin for feedback, and our partners at the Massachusetts Department of Elementary and Secondary Education as well as the Lawrence Public Schools, Holyoke Public Schools, Southbridge Public Schools, and the Springfield Empowerment Zone Partnership.

Motivation

Educational inequality based on race, ethnicity, and socioeconomic status remains unacceptably high in the United States. This was true prior to the COVID-19 pandemic for a variety of important measures of academic achievement and educational attainment, including standardized test scores (Reardon, 2011; Carter & Reardon, 2014; Hashim et al., 2020; Hanushek et al., 2020), years of schooling (Duncan & Murnane, 2011), and college going (Bailey & Dynarski, 2011). Unfortunately, there are signs that these patterns of inequality have only been exacerbated due to the pandemic's disproportionate impacts on low-income communities and communities of color (e.g., Kuhfeld, Soland, Tarasawa, Johnson, Ruzek & Liu, 2020).

One way policymakers have attempted to confront such inequality is through "turnaround" reforms, which represent efforts to rapidly improve outcomes for public K-12 schools identified as among the lowest-performing in a given state. These schools disproportionately serve low-income children and students of color, and therefore, significantly improving these schools could put a meaningful dent in educational inequity. Significant federal funding has been devoted to these turnaround efforts, including the Obama Administration's $7 billion School Improvement Grant (SIG) program. These efforts continue today in a somewhat different form under the federal Comprehensive Support and Improvement (CSI) program which provides more leeway for local leaders to design policy responses for low-performing schools than the Obama-era programs (Meyers et al., 2022). This motivates the continued need for research on best practices in school improvement policy to inform these state and local decisions.

The existing turnaround literature has focused primarily on school-level improvement efforts; however, there are theoretical and empirical reasons to believe that more attention should be paid to district-level turnaround reforms. Scholars from organizational behavior, economics, and political science traditions alike have made the theoretical case that school districts play a crucial role in the education production function (Blazar & Schueler, 2022). Scholars of education have argued that districts may have greater capacity to create conditions for schools to succeed than individual schools on their own (Supovitz, 2006; Zavadsky, 2013; Johnson et al., 2015). Quantitative evidence decomposing variation in student achievement illustrates that districts play a non-trivial role in producing these outcomes (Chingos, Whitehurst & Gallaher, 2015). In the context of turnaround, a recent meta-analysis found suggestive evidence that districtwide improvement efforts were associated with greater gains on average than school-level turnaround reforms (Schueler et. al., 2021). Indeed, a recent report from a long list of prominent scholars and leaders, sponsored by Brown University's Annenberg Institute highlights the pressing need for new research focused on district leadership and policy (Schwartz et al., 2023).

The research base on districtwide reforms designed to improve low-performing school systems is made up of a small number of case studies that demonstrate variation in the impacts of these reforms. For example, Harris and Larsen's (2022) work on educational outcomes in New Orleans shows that the reforms enacted in the aftermath of the Hurricane Katrina disaster and the state takeover of the district that followed generated notable improvements in student achievement outcomes. Schueler, Goodman, and Deming (2017) also find that state takeover and turnaround in Lawrence, Massachusetts generated positive achievement gains in the early years of reform; however, researchers have documented more mixed or even negative results in contexts such as Newark, New Jersey (Chin et al., 2019), Philadelphia, Pennsylvania (Gill, Zimmer, Christman & Blanc, 2007), and Tennessee (Zimmer, Kho, Henry & Viano, 2015).

One approach to district-wide turnaround that has become increasingly common over time is state takeover, which represents a change in school governance that typically removes authority from a locally elected school board and places decision-making power with the state. Studies from the pre-No Child Left Behind (NCLB) era found that states had not had much success at improving academic outcomes via takeover (Wong & Shen, 2002, 2003). More recent research on the impact of takeovers on test-score outcomes shows that these reforms yielded no academic benefits, on average, for the targeted districts and provides suggestive evidence that takeover can be disruptive for reading achievement in the early years of reform (Schueler & Bleiberg, 2022). Although takeover effects are null on average, scholars found significant heterogeneity of effects across districts (some positive, others negative). Researchers observed variation in impacts both across and within states, suggesting that state capacity does not explain all of the variation in effects. In other words, it does not seem to simply be the case that some states are better at implementing takeover than others. This motivates the need for cross-district case studies within states to begin to reveal what makes some takeover efforts more successful at enhancing academic achievement than others.

Another limitation of the existing research on school and district turnaround is that it has tended to focus on the short-term impacts of reform with limited attention to whether initial impacts are sustained over time. A meta-analysis of evaluations of post-NCLB turnaround interventions found that two-thirds of estimates examined less than three years of reform implementation (Schueler et al., 2021). This is an unfortunate omission because the evidence prior to NCLB concluded that school and district improvement takes time. More specifically, typically at least three years of reform implementation is necessary for these interventions to demonstrate results (Desimone, 2002; Borman et al., 2003; Gross et al., 2009; Peurach & Neumerski, 2015). The few studies of post-NCLB district improvement that do examine medium-term outcomes again yield mixed results (Harris & Larsen, 2022, Pham et al., 2020; Chin et al., 2019). Therefore, more research is needed on the effects of these types of reforms beyond the first two years.

Additionally, there has not yet been significant attention to the issue of whether districtwide reforms are replicable across contexts. The case studies mentioned above tend to focus on cases that were exceptional in some way and the first of their kind in their respective contexts. For example, the New Orleans reforms were implemented in the context of a major national disaster and resulted in what is now the only all-charter school district in the country. The Lawrence turnaround represented the first takeover that the state of Massachusetts had enacted after the passage of the state's 2010 Achievement Gap Act which gave the state greater authority when intervening in low-performing systems. Therefore, it is important to understand whether it is possible to transport effective districtwide reform efforts and successfully apply these practices in new contexts. There is also an open question of whether states have capacity to effectively support more than one takeover at a time, as they try to replicate early successes in new contexts. On the one hand, it could be difficult for state agencies to provide support in multiple districts at one time. On the other hand, it is possible that states could learn from earlier interventions and get more effective at supporting improvement over time.

Furthermore, past turnaround evaluations have focused primarily on test score outcomes, despite non-test outcomes' predictive power for students' later success. In particular, a growing evidence base shows that non-test score outcomes like attendance or suspensions are also important for long-term outcomes (e.g., Heckman, Stixrud, and Urzua, 2006). In fact, Jackson (2018) finds that teacher value-added to these types of outcomes are much more predictive of

long-term success that their value-added to test-score impacts alone. Additionally, it is particularly important to examine non-test outcomes when evaluating test-based accountability policies because it is possible that these programs can lead test scores to become overemphasized at the expense of other important non-test outcomes (Deming et al., 2016). While there is suggestive evidence that turnarounds do not appear to hurt non-test score outcomes on average, there have been very few studies that have examined the impacts of these reforms on non-test outcomes—eight on the most studied non-test outcome according to a recent meta-analysis (Schueler et al., 2021; for a more recent exception see Eren et al., 2023). There is an additional need to understand the mechanisms through which district turnaround succeeds or fails, and it is likely that human capital plays a major role, but few studies have examined turnaround impacts on the composition of the teaching forces in these contexts (for exceptions see Lincove et al., 2018 and Pham et al., 2020). Therefore, the potential for important non-test score outcomes to be inadvertently harmed—or improved—as a result of turnaround and the need to understand the mechanisms of improvement motivates a need to look beyond test score outcomes.

Massachusetts is a valuable context for studying the sustainability and replicability of districtwide turnaround reforms on test and non-test outcomes. The state is home to the Lawrence Public Schools, a historically low-performing district that was taken over by the state in 2012. Previous research documents that the reforms implemented by state-appointed leaders resulted in meaningful academic improvements (particularly in math achievement) in the first two years of implementation, making Lawrence a rare positive proof point for the improvement of a persistently low-achieving district serving a large concentration of low-income students of color (Schueler, Goodman & Deming, 2017). Based at least in part on this success, the state has continued to be involved with district improvement efforts in Lawrence and has also intervened in three other low-performing contexts—Holyoke, Southbridge, and Springfield—all of which serve large numbers of low-income children of color.

More specifically, Massachusetts enacted state takeovers in Holyoke and Southbridge. Springfield avoided the threat of state takeover by partnering with the state on a novel form of state-led turnaround in which an independent board made up of state appointees and local representatives oversees a set of the district's low-performing schools in what is called the Springfield Empowerment Zone Partnership (SEZP). The non-profit organization supporting SEZP has now led the creation of similar "empowerment zones" based on the SEZP model in ten different states across the country (Empower Schools, 2023). This model has not yet been subject to rigorous independent evaluation to our knowledge. Given the need for more evidence on the sustainability and replicability of district-wide turnaround efforts, we address the following research questions in the context of Massachusetts:

(1) Was the state able to sustain the early Lawrence achievement gains?

(2) Was the state able to replicate the early Lawrence gains in other low-performing contexts (Holyoke, Springfield, and Southbridge)?

(3) How did these state-initiated reforms impact the characteristics and stability of the teaching force in each context?

State-Initiated District Turnaround in Massachusetts

Massachusetts is a relatively high-performing state when it comes to K-12 education but struggles with persistent inequality of educational outcomes (Papay et al., 2020). In addition to its school-level improvement efforts, the state has also been engaged in a number of initiatives

targeting entire low-performing districts or large clusters of schools within districts. This was, in part, made possible by the passage of the 2010 Achievement Gap Act (AGA) which allowed for state takeover (or "receivership") of entire low-performing districts. Once placed in receivership, the state's Commissioner of Elementary and Secondary Education appoints a "Receiver" who assumes all the decision-making power previously held by the superintendent as well as the elected school board. The Receiver then enjoys broad authority to make district-wide policy changes and even has the ability to limit, suspend or change provisions of the existing collective bargaining agreement and require all staff to reapply for their positions (Commonwealth of Massachusetts, 2010). The state legislature has recently begun considering a new bill called the "Thrive Act" which would, among other things, reduce the power given to the state by the AGA to take over low-performing districts.

The state enacted its first takeover after the passage of the AGA—of the Lawrence Public Schools—in 2012 under the leadership of then Commissioner Mitchell Chester. Previous research documented that the Lawrence turnaround generated substantial gains in math in its first two years without slippage on other indicators (Schueler, Goodman & Deming, 2017). In part based on those early successes, the state maintained a significant leadership role in Lawrence, undertook takeover of two additional districts—Holyoke and Southbridge—and embarked on a unique state-initiated governance arrangement in a zone of schools within the Springfield Public Schools called the Springfield Empowerment Zone Partnership (SEZP). SEZP is the only context under study here that is not a formal example of state takeover or receivership.

The Massachusetts Department of Elementary and Secondary Education (MA DESE) provided technical support to takeover districts throughout this period and established a new office in 2016 to focus more specifically on supporting receivership contexts in a variety of areas ranging from operations to academic improvement. This office assesses local conditions, helps identify receiver candidates for the Commissioner, represents the Commissioner during collective bargaining in the receivership districts, helps develop the overall improvement strategy within each context, provides ongoing support to the Receivers, and strategizes regarding transitions out of receivership. Up until this point, the State had enjoyed leadership stability—Chester was the longest serving chief state school officer in the country when he passed away unexpectedly in the summer of 2017. In April of 2018, Commissioner Chester was succeeded by Jeffrey Riley who had up until then been serving as the state-appointed Receiver in Lawrence. See Figure 1 for a timeline of leadership transitions.

As we show in Table 1, all four of the contexts targeted for turnaround were performing well below not only the statewide average (by between -1.20 and -0.60 standard deviations on math assessments) but also the average for majority low-income districts on standardized tests prior to the reforms. They all served majority low-income student populations with high concentrations of students of color. As we show in Table 2, all four contexts had higher concentrations of first-year teachers than the rest of the state leading up to reforms. They also each had larger shares of Hispanic teachers, though none came close to having a teaching force that was demographically reflective of the student populations as all had majority-white teaching populations (between 68 percent White for SEZP Cohort 2 and 93 percent for Southbridge).

In addition, there were similarities in the reforms pursued in each place. For example, all districts implemented new teacher compensation systems that included a career ladder, a pay-scale based in part on performance, stipends for extended learning time and serving in teacher leadership roles, as well as pay increases. This was the result of negotiations with the unions in all four contexts that resulted in the ratification of new collective bargaining agreements. This

was a particularly interesting choice given Receivers were not required by the State law to do so. In all four contexts, leaders also prioritized the diversification of the educator workforce. However, there were also notable differences between the four contexts and the policies leaders pursued in each place that we describe in the next section. Throughout this paper we focus on reforms undertaken after the AGA passage but prior to the onset of the COVID-19 pandemic. We focus on the pre-pandemic period given COVID-19's impacts differentially affected turnaround schools and communities (Harbatkin, McIlwain & Strunk, 2022). One implication of this decision is that we are only able to examine a few post-reform years of results for some of the reform contexts. For example, we only observe three years of post-takeover outcome data for Southbridge, although it is possible that the results in more recent years may have changed. We refer to academic years with the spring year (e.g., 2015-16 is "2016").

Lawrence. Lawrence is a mid-sized, post-industrial city about 40 minutes north of Boston by car. The district serves a student population of nearly 13,000 students in roughly 30 schools. Almost all students are growing up in low-income homes (92 percent), as we show in Table 1. Prior to turnaround, 88 percent of the student population was Hispanic, and 82 percent had a first language other than English. Lawrence is home to large communities of families who recently arrived in Massachusetts from Puerto Rico or the Dominican Republic. Based on persistent low performance as well as leadership challenges, Massachusetts placed the district into receivership and appointed a Receiver who began implementing turnaround efforts in the 2012-13 year. At the time, the district performed -0.28 standard deviations (SD) below the national mean on ELA exams and -0.20 on math based on the Stanford Education Data Archive (SEDA) which allows for achievement comparisons across states by norming state exams to the National Assessment of Educational Progress.

The reforms were characterized by a focus on increasing school-level autonomy—at differentiated levels depending on school performance and perceived capacity—and holding schools to higher expectations. The central office budget was reduced by 25 percent and funds were pushed to the school level. As we show in Figure A1, the state reports that per pupil spending did not increase in Lawrence relative to the increases statewide. If anything, in the more recent years of reform, it declined. Principals, alongside teacher leader teams, were given autonomy over their calendars, interim assessments, staffing, and more. Most of the schools remained under district management but a small number were handed over to outside operators ranging from a charter management group, a local non-profit, and the local teachers' union. All schools retained neighborhood-based student assignment and a unionized teaching force. In year four of takeover, leaders embarked on a high school redesign process which was again revamped in year seven.

Throughout the period we study, Lawrence leaders prioritized increased learning time—extending the school day and/or year, building out extracurricular options in collaboration with community partners, and offering tutoring for students in need of support. Notably, the district ran "vacation academy" programs (called "acceleration academies" in Lawrence) for students below proficiency thresholds on standardized exams. For these programs, the district recruited teachers they considered to be particularly effective from within and beyond the district to come work with small groups of about ten students in a single subject—most often math or ELA—over a week-long vacation break. Previous work on the first two years of Lawrence reforms shows participation in these academies explained roughly half of the post-turnaround gains in math and all of the gains in ELA (Schueler, Goodman & Deming, 2017).

Another focus of the reforms was on improving human capital throughout the district. The Receiver's team was particularly aggressive when it came to school leaders, replacing half of all principals by year two of the reforms. They actively replaced a smaller share—roughly ten percent—of all teachers in those early years, although more left voluntarily (Moore Johnson, 2017). The reforms also placed an emphasis on using data to drive instructional improvements, and in the later years on shifting all schools to vetted, standards-aligned curricula, building out early college programs, and enhancing family engagement.

Starting in the 2017-18 year, as part of efforts to begin a process of returning local control, the State appointed a Board—the Lawrence Alliance for Education (LAE)—to serve as Receiver, oversee the Superintendent, and include local leaders in decision-making. After that year, the original Receiver left the district to become State Commissioner and the LAE Board hired a new Superintendent (Moore Johnson, 2021). Weeks into the new Superintendent's tenure, Lawrence experienced two major gas explosions, killing one former Lawrence student, leaving many families displaced from their homes, and leading to school evacuations for suspected gas leaks in the fall of 2018. We raise this because we later explore whether declines in outcomes appear due to the effects of these tragic events.

In the year following the gas explosions, the new leaders pushed forward with reforms such as developing a more explicit performance management framework to set common expectations for schools about how to earn autonomies and how school-level funding operated, creating structures for principal collaboration, hiring a new principal who worked on increasing coherence across programs at the high school, standardizing the calendar districtwide, and attempting to build support for a restorative justice approach to discipline. These new reforms were just getting underway in the year prior to the onset of the COVID pandemic. See Appendix Figure A2 for a summary of the reforms over the seven-year period we study here.

Holyoke. Holyoke—the context for Massachusetts' second takeover under the AGA—is a small city in Western Massachusetts about an hour and forty minutes from Boston by car. The district serves nearly 5,000 students in roughly 12 schools. As we report in Table 1, a large majority of students were growing up in low-income homes (84 percent) and were identified as Hispanic (77 percent). A large share—but smaller share than in Lawrence—had a first language other than English (59 percent) and again a large but smaller share than Lawrence was born outside of the continental U.S. In March 2015, three years after the Lawrence takeover, Commissioner Chester recommended the state takeover of Holyoke Public Schools (HPS) and appointed a Receiver who began implementing reforms starting in the 2015-16 year and who was at the helm of the district for the entire period under study here. At the time, the district was performing -0.58 SD below the national average on ELA performance and -0.38 in math, again based on the SEDA data. The community in Holyoke appeared to express greater resistance to the takeover in the early years of reform than was observed in Lawrence (Schueler, 2019).

The Holyoke reforms also involved increasing school-level autonomy, though in large part due to the district's smaller size, some things remained more standardized across schools than they had in Lawrence, such as the calendar. The Receiver hired a new central office cabinet and built out a team to directly support principals. The reforms in the years under study tended to target the district's youngest and oldest children. For example, the team expanded pre-Kindergarten programs significantly. The leaders also redesigned the two high schools into a single campus, invested in career and technical programs, created a menu of pathway programs, created a robust early college program, and increased the availability of advanced coursework for

high school students. Later, starting in the 2018-19 year, the district handed over management of one middle school to an independent charter operator (that remained a traditional public school).

The Holyoke reforms also included extended learning time, vacation academies, enhanced enrichment offerings, efforts to improve human capital including principal and teacher replacements—similar to Lawrence with a heavier emphasis on replacing school leaders rather than teachers—using data to drive instructional improvement, engaging families, and addressing deferred maintenance to facilities as well as basic operational systems like a phone communication solution for contacting families districtwide. There were increased efforts to ensure students with disabilities were being served in the least restrictive environment possible and to incorporate more feedback from families of students with disabilities through the creation of the Holyoke Special Education Parent Advisory Council. These moves occurred in the aftermath of pre-takeover allegations of physical abuse of students with disabilities in one particular intervention program. Leaders also expanded the dual language program and began introducing new curricular materials although not consistently districtwide until the later years of reform outside the window of our study. These changes are summarized in Appendix Figure A3. Funding increases throughout the reform period did not outpace statewide increases. If anything, per pupil spending declined in Holyoke post-takeover somewhat relative to the rate of change in the rest of the state (see Appendix Figure A1).

Springfield Empowerment Zone Partnership. Springfield is a medium-sized city—larger than Lawrence but smaller than Boston—in Western Massachusetts, just a 15-minute drive south of Holyoke. Leading up to reforms, nearly all students were growing up in low-income families (90 percent). A majority of students identified as Hispanic (52 percent). Springfield served a larger share of Black students (20 percent) than any of the other contexts under study and a smaller share of students whose first language was not English (28 percent) than both Lawrence and Holyoke. Leading up to intervention, the district as a whole was performing -0.32 SD below the national average on ELA tests and -0.26 in math. Under the threat of receivership due to persistently low academic performance, the state and district agreed to a new, unique model for school improvement that allowed the district to avoid state takeover but still undertake state-initiated improvement efforts. Specifically, six middle schools (including one school serving grades 6-12), serving roughly 4,000 students, categorized as "underperforming" in the State's accountability system were placed in a "zone" and targeted for improvement through what was named the Springfield Empowerment Zone Partnership (SEZP) (we call this "SEZP Cohort 1").

An MOU was signed between MA DESE and the Springfield Public Schools indicating that the SEZP would be governed by a non-profit Board of Directors. This Board is made up of the Mayor, the Superintendent of Springfield Public Schools, the Vice Chair of the School Board, and four State Commissioner appointees who are based in the region. Therefore, the State appoints a majority of Board members, but the Board is intended to also provide greater local representation and influence than what would exist under a typical formal state takeover (Jochim & Opalka, 2017). The majority of Commissioner-appointed Board members have also had local Springfield ties (e.g., a minister, a family foundation officer, a non-profit leader). There was relative stability of leadership on the Board over the full period under study, including the same Board Chair and Superintendent. SEZP was incubated by a separate non-profit organization, Empower Schools, which is led by several people who were involved in shaping the Lawrence reforms in its early years, and which has now supported the creation of "empowerment zones"

based on the SEZP model in ten different states (Empower Schools, 2023). The first year of reform was 2015-16, the same year that the Holyoke takeover began.

The SEZP reforms extended the approach taken in the early years of the Lawrence turnaround, by granting school-based autonomy in exchange for a heightened level of accountability. Although unlike in Lawrence, the same level of autonomy was granted across all SEZP schools from the start. Principals and their teacher leader teams had the authority to make decisions related to budget, curriculum, staffing, schedule, and school culture (up to 80 percent of the budget) and the district provided a menu of services that schools could select (or not). In the early years, three of the middle schools were reconfigured such that the Zone included a total of nine distinct learning communities. One of the new schools was managed by a charter operator, but none were converted to charter status, and all remained unionized. In 2017-18, a large high school serving 1,400 students was added to the Zone and reconfigured into two new learning communities over two years (we call this "SEZP Cohort 2").

From the start of SEZP, learning time was expanded across all Zone schools, tutoring offerings were expanded, and vacation academies were provided to students struggling to meet proficiency benchmarks. A field experimental study shows that these week-long programs improved test scores and reduced exposure to exclusionary discipline for participating students (Schueler, 2018). There was again an emphasis on replacing school leaders and, to a lesser extent, teachers. SEZP emphasized the use of data for planning, accountability, and instructional improvement. Leaders also established new dual language and early college programs. These changes are summarized in Appendix Figure A4. While per pupil spending in Springfield outpaced the state pre-reform, the funding post-takeover declined relative to the state, though we can only examine spending for the full district, not SEZP specifically (See Figure A1).

Southbridge. Southbridge is the final context and the most recent Massachusetts district to enter receivership. It is a small city that is about a 75-minute drive to the southwest of Boston. In the period of study, the district served roughly 1,800 students total in six schools. As we report in Table 1, the district had the lowest share— but still a large majority—of low-income students (73 percent) out of the four contexts. It also had the largest share of White students (48 percent) but still served a sizeable share of Hispanic students (47 percent). About one-third of the students had a first language other than English. After placing the district into receivership, the Commissioner appointed a Receiver who began in 2016-17. At the time, the district was performing -0.30 SD below the national average in ELA and -0.41 in math.

The Southbridge turnaround was marked by leadership instability. The first Receiver was placed on administrative leave after her first year and replaced for the first half of the 2017-18 year by a state-level leader who served as Interim Receiver until a more permanent Receiver was appointed midway through the 2017-18 year. This Receiver remained through the rest of the period under study (and beyond). Southbridge was no stranger to leadership churn as, prior to state takeover, the district had seven superintendents and seven principals at its high school over the previous six years. The State agency was supporting the district while also supporting two other receivership districts (Lawrence and Holyoke) at the same time.

The first year of reform in Southbridge was focused on increasing alignment across schools. The State hoped to eventually increase school-level autonomy but did not see this as a possibility at the outset given perceived school capacity limitations. The Receiver established a new alternative high school program for students who had behavioral issues, extended learning

time for elementary school students, focused on principal and teacher replacements, and negotiated a new contract (modeled on those from the other three contexts). The second year was focused on stabilization given the leadership transition. Schools began shifting to vetted, standards-aligned curricula across the district, added family liaisons at each school, and established new translation and interpretation services for families.

In the third year of reform, the new Receiver focused on creating structures for principal collaboration and capacity building, redesigning the alternative high school into a therapeutic day school, adding time for teacher professional development and planning, using data to drive improvements, and implementing the Positive Behavior Intervention and Supports (PBIS) tiered framework for improving student behavior. In this year, leaders also shifted from a paper-based record-keeping system to a digital record-keeping system for student information management, finance, human resources, facilities, operations, budget, food service, and more. These new reforms were just getting underway in the year prior to the onset of the COVID pandemic. We further summarize the policy changes in Appendix Figure A5. Unlike the other turnaround contexts, the state reports that per-pupil spending increased in Southbridge in the post-takeover period relative to the rest of the state (see Figure A1).

Data

To assess the impact of these reforms on student outcomes, we leverage statewide, longitudinal, student-level data provided by MA DESE for the school years of 2006-07 to 2018-19 (the last full pre-COVID year). These data include each student's grade, school, district, demographic characteristics, standardized test scores, attendance, and discipline record by year. The data includes more than 500,000 unique student observations per year. Our preferred analytic sample includes roughly 25 percent of the full universe of Massachusetts students who are within the 54 districts that served a majority low-income student population in the pre-treatment period. This is a more relevant set of comparison districts given that all treated districts are majority low-income and due to the well-established correlation between socioeconomic status and academic outcomes.

The outcomes of interest consist of students' academic performance as measured by their test scores on the statewide math and English Language Arts (ELA) exams, administered annually in grades 3-8 and grade 10, as well as science exams administered annually in grades 5, 8, and 9 or 10. We standardize these scores within year, subject, grade, exam, and modality (computer vs. paper) using the full sample of Massachusetts students. Standardizing within exam is necessary because there was variation over time and within years, with all students taking the Massachusetts Comprehensive Assessment System (MCAS) in years prior to 2015, some students taking the Partnership for Assessment of Readiness for College and Careers (PARCC) exam in 2015 and/or 2016, and then all students switching to the MCAS 2.0 in 2017 and beyond. Additionally, in 2015 and 2016, about half of the students who took the PARCC also took Computer-Based Testing (CBT) while the other half took paper exams. From 2017 to 2019, an increasing share of students took CBT (Backes & Cowan, 2019). We also confirm prior to standardizing the test scores (examining the raw scaled scores) that there was not a substantial change in the presence of floor or ceiling effects when these exam shifts occurred either in the treated or comparison districts that could have artificially resulted in perceived gains or losses for the treatment groups relative to the comparison group (see Appendix Figure A6).

We also examine non-test measures including the number of days a student is marked as having attended school within a given year, the number of in-school suspensions and separately the number of out-of-school suspensions the student received, and finally whether a student was retained and progressed to the next grade in the next year. To assess the impact of reforms on the teaching force, we leverage statewide longitudinal, teacher-level data from MA DESE from 2007-08 to 2018-19. These data include demographic information on each teacher's gender and race/ethnicity. We also observe a hire date and use this to calculate a proxy measure for years of experience that is based on the year the staff member was first hired by the public school system in Massachusetts (regardless of whether that person was hired as a teacher or a different type of staff member). We generate a variable for whether a teacher was in the first year we observe them as a teacher in the state of Massachusetts as another proxy for experience level. Finally, we calculate two measures of teacher turnover—an indicator for whether a teacher left teaching and/or left the state of Massachusetts and another indicator for whether the teacher transferred to a new school district within the state. For both SEZP cohorts we code the non-SEZP Springfield Public Schools as a separate district from SEZP to estimate between-district movement (i.e., if a teacher transferred from SEZP to a non-SEZP school in Springfield, we consider that a between-district transfer here).

Analytic Methods

Examining Impacts on Student Outcomes. To study the effect of turnaround, we conduct difference-in-differences analyses that compare achievement trends of students in turnaround contexts to achievement trends of students in comparison districts that did not experience state-led turnaround. Because the reforms varied fairly substantially between districts, we estimate turnaround impacts separately for each context (excluding the other ever-treated districts from the sample) rather than estimating a staggered difference-in-difference model combining all treated districts together. For SEZP, we examine two cohorts separately, as the reforms began in 2016 (Cohort 1) but a new high school was added to the zone starting in 2018 (Cohort 2). We begin by running event study models to transparently assess the parallel trends assumption and to examine how effects may have developed over time. In all student-level models, we treat the six years leading up to turnaround as the pre-treatment period and omit the last year prior to the intervention as the comparison year. Our primary specification is a school-by-grade fixed effects model as follows (using Lawrence as an example):

$$Y_{isgy} = \beta_0 + \beta_j \sum_{j=2007}^{2019} Lawrence_{isgy} \times Year_y + \theta_{sg} + \gamma_{gy} + X'_{isgy} + \varepsilon_{isgy} \quad (1)$$

Here, *Y* is an outcome, such as a standardized math test score for student *i* in school *s* and grade *g* in year *y*. $\sum_{2007}^{2019} Lawrence_{isgy} \times Year_y$ represents a series of interactions between a binary indicator for whether a student was enrolled in the Lawrence Public Schools and a binary indicator for whether it was a given year. These interactions in the post-period provide estimates of the extent to which changes in Lawrence's outcomes in the years after the turnaround reforms differ from changes in the comparison districts over the same period. These changes are in reference to the omitted year immediately preceding the start of the reforms (which in the case of Lawrence is 2012).

We include school-by-grade fixed effects (θ_{sg}) to generate estimates comparing the same school-grade combinations to themselves over time. These fixed effects also provide the main

effect of ever being treated for the purpose of the difference-in-difference estimates. Grade-by-year fixed effects (γ_{gy}) control for any shocks – unrelated to turnaround reforms – that would have impacted achievement for students in a particular grade in a specific year such as changes in exam difficulty. We include student-level covariates (X') to account for compositional changes within either the treated or comparison districts over time. These include binary indicators for whether a student was identified as female, Black, Hispanic, Asian, low-income, and an immigrant, as well as whether the student had a first language other than English. These covariates help address any compositional shifts in the treated or comparison district student populations that could have occurred over time based on observable student characteristics, although we show in Appendix Table A1 that any such changes were relatively minor.

Given previous research showing that changes to exam and modality in Massachusetts impacted student test performance, particularly among students receiving special education services and whose first language was not English (Backes & Cowan, 2019), we include a set of test-related controls when estimating impacts on test-based achievement in math, ELA, and science. Specifically, we control for whether the student took the PARCC exam, whether a student took a computerized exam, whether the test modality was new to the student in that year, and interaction terms that allow the impact of a computerized exam to vary for students identified as special education or having a first language other than English. We cluster standard errors at the district level.

We also examine whether the impact of reforms on student outcomes varied depending on student demographic characteristics. There are two differences between the models we use for these analyses and model (1). First, instead of the event study models which estimate separate coefficients for each year, we run basic difference-in-difference models where we pool all pre-treatment annual effects and all post-treatment annual effects together. Second, we interact the post-treatment indicator with the student demographic characteristic to test whether the treatment effect varied for a particular subgroup. For example, we interact an indicator for whether the student was identified as Black with the post-treatment indicator to examine whether treatment effects were different for Black students, on average. We do this separately for each demographic characteristic and each treatment context.

Synthetic Control Methods Examining Impacts on Student Outcomes. The key underlying assumption of the difference-in-differences approach is that the comparison group and treatment group were on a similar trajectory on the outcome prior to the intervention. However, for some of the treated contexts and student outcomes, we observe violations of this parallel trends assumption using model (1). As a check on whether any findings are driven by differences between the treatment and comparison groups in pre-treatment trends, we also use the synthetic control group method to identify comparison groups that were on a similar trajectory with respect to the outcome leading up to the reform implementation (Abadie, Diamond & Hainmueller, 2010; Abadie, 2021). We use the method to identify the weighted combination of all other untreated majority-low-income districts in Massachusetts that minimizes the mean squared prediction errors of the outcome variable of the treated district in each of the pre-treatment years. We exclude districts not observed in every year to create a balanced panel for the synthetic control package. For SEZP Cohort 2, which consisted of a single high school, we use other high schools to make up our donor pool rather than districts. We then generate difference-in-differences estimates using the synthetic control as the comparison group.

It is not possible to use traditional statistical inference approaches to infer the statistical significance of results in a synthetic control group framework because doing so typically involves analyzing the data at the level of assignment to treatment (in this case, typically the district level) dramatically reducing the sample size. Instead, we follow Abadie, Diamond and Hainmueller (2015), Hernandez (2019), and McClelland & Gault (2017) to conduct "placebo studies" based on the idea that we would not expect to observe estimated effects similar or greater in magnitude to those for the treatment groups in districts where the reforms did not occur. To do this, we temporarily assign treatment status to each placebo district in the donor pool, and then conduct the synthetic control group analysis generating estimate effects. Finally, we compare the treatment effects for our treated contexts to the distribution of estimated placebo effects. Where a majority of placebo effects are smaller in magnitude than the treatment effects, we have greater confidence in the estimated treatment effect (Billmeier & Nannicini, 2013; Shores et al., 2022).

Examining Impacts on Teacher Outcomes. To examine whether the reforms impacted characteristics of the teaching force, we estimate models of the following form (again using Lawrence as an example):

$$Y_{tsy} = \beta_0 + \beta_j \sum_{j=2008}^{2019} Lawrence_{tsy} \times Year_y + \theta_s + \gamma_y + X'_{tsy} + \varepsilon_{tsy} \quad (2)$$

Here, *Y* is a binary outcome, such as an indicator for whether a teacher was identified as Black for teacher *t* in school *s* and in year *y*. $\sum_{2008}^{2019} Lawrence_{isy} \times Year_y$ represents a series of interactions between a binary indicator for whether a teacher was working in the Lawrence Public Schools and a binary indicator for whether it was a given year. These interactions in the post-period provide estimates of the extent to which changes in Lawrence's outcomes in the years after the turnaround reforms differ from changes in the comparison districts over the same period. These changes are in reference to the omitted year immediately preceding the start of the reforms (which in the case of Lawrence is 2012). For teacher outcomes, the first pre-treatement year we estimate for Lawrence is 2008 because our data do not allow us to go back quite as far for teacher outcomes as student outcomes.

We include school fixed effects (θ_s) to generate estimates comparing teachers in the same schools to themselves over time. These fixed effects also provide the main effect of ever being treated for the purpose of the difference-in-difference estimates. Year fixed effects (γ_y) control for any shocks – unrelated to turnaround reforms – that would have impacted achievement for students (and provide the main effect of post-treatment). We include teacher-level covariates (*X'*) including binary indicators for whether a teacher was identified as female, Black, Hispanic, Asian, and White (except we exclude race/ethnicity controls when examining these as outcomes). We also control for a teacher's years of experience and whether they were a first year teacher (omitting these when examining experience variables as outcomes). Our results are not sensitive to the inclusion of these covariates. Like with the student outcomes, we also calculate pooled difference-in-difference estimates where, unlike the event study models, we pool all pre-treatment years and pool all post-treatment years to get a summative sense of the reform impacts over the entire post-treatment period. Again, for all teacher models we cluster at the district level.

Findings

Lawrence. Overall, the results of the Lawrence reforms were positive to neutral. We begin by displaying the results for test score outcomes graphically in Figure 2. This is part of a series of figures that all provide descriptive outcome trends for the treated context and the comparison districts in the left-most panel, regression-based estimates of the effects in the second figure from the left, synthetic control-based estimates in the second figure from the right and results from the placebo tests in the right-most panel. For Lawrence, we observe large positive impacts of the reforms on math achievement, that increased in magnitude for the first three years of the reforms, level off for the fourth and fifth year, and began to decrease in the last two years of the reforms, coinciding with the shift to an appointed board (2018) and the arrival of a new superintendent (2019). That said, the impacts remain positive even in year seven. The average effect across all seven post-takeover years was 0.21 standard deviations (SD), as we show in Table 4. Again, this is combining effects across all tested grades.

In ELA, the positive effects in the early years of reform were more modest in magnitude and began to trend downward in the last two years of reform that we observe. The overall pooled effect was not statistically different from zero (see Table 4). The declines in the last reform year we observe (2019), which coincided with the gas explosions in the Lawrence community, do not appear to be due to these events alone, as the declines persist even after we exclude the schools located in neighborhoods most directly affected by the explosions (see Appendix Figure A8). In science, positive effects began to emerge in the second year of the reforms and increased in magnitude until they began to dip, although remained positive, in the last year that we observe. The pooled impact across all post-takeover years was 0.12 SD (see Table 4). Based on our visual inspection and joint F-tests of the pre-treatment effects, reported in Table 3, none of the test score impacts appear to be driven by differences in pre-trends between Lawrence and the other majority low-income districts. We confirm results do not seem to be the result of increased rates of missingness on the outcome measures post-treatment in Appendix Table A2.

Turning to non-test outcomes, the Lawrence reforms appeared to increase the average number of days students attended school by 1.78 days pooling across all post-takeover years (Table 4). However, the F-test of whether the pre-treatment coefficients are jointly statistically significant (reported in Table 3) and a visual inspection suggests that these attendance impacts may have been driven by pre-trends. The effects are, however, robust to the use of synthetic control methods (shown in Figure 3), allowing us to compare Lawrence to a synthetic comparison district with a very similar attendance pre-trend. In the upper right-hand corner of Figure 3, we show that this does not appear to be due to chance as the majority of placebo district effects were smaller in magnitude than the effects we observe for Lawrence.

When it came to disciplinary outcomes, the reforms appeared to decrease both in-school suspensions by 0.06 suspensions as well as out-of-school suspension by 0.03 suspensions after pooling effects across all post-takeover years (see Table 4). There is some evidence that the in-school suspension result could have been driven by pre-trends, but the effects are robust to the use of synthetic control methods, as we show in Figure 3. For out of school suspensions, the decreases were larger in the early years of reform and became neutral from 2016 to 2019. Finally, the Lawrence reforms also appeared to increase the rate of grade progression, but we cannot rule out that this result may have been driven by pre-trend differences and the results were not robust to the use of synthetic control methods. As we show in Appendix Figure A7, none of these results are sensitive to the inclusion of student fixed effects—allowing us to compare students to themselves over time—suggesting that compositional shifts in the student population in or out of Lawrence do not drive the findings.

In Table 4, we display results for an examination of whether the effects of the Lawrence reforms varied for students based on their demographic characteristics. For most outcomes, we find that the positive impacts were largest for students of color and low-income students with the exception of discipline outcomes for which the effects were somewhat smaller for these groups, although still positive. In contrast, the test score impacts were smaller for students identified as immigrants. In Table 5 we demonstrate that the gains for most outcomes were larger for middle school and high school students than elementary schoolers. There were two exceptions—both the attendance and grade progression impacts were almost entirely concentrated among high school students.

Next, we examine whether and how the Lawrence takeover impacted the characteristics of the teaching force in Lawrence, particularly given the various reforms focused on improving human capital including the ratification of new collective bargaining agreements across all four contexts. We display the event study results visually in Figure 4. We find that the reforms increased the levels of annual teacher turnover, driven both by an increase in the share of Lawrence teachers leaving teaching and/or the state and increasing in the share transferring out of the district. Pooling all post-takeover years, the reforms increased the share leaving teaching in the state by three percentage points and between-district transfer by six percentage points (see Appendix Table A3). Increases in the rates of teacher transfer and departure persisted throughout the whole seven-year post-treatment period. In other words, this was not a temporary increase. The teaching force after the takeover also appears to have become less experienced overall featuring a four percentage point higher share of first year teachers, although we cannot fully rule out the possibility that this was because the treatment group was on a somewhat different trajectory than the comparison group prior to the takeover. Finally, after the reforms, Lawrence was home to a slightly greater share of Hispanic teachers in five of the seven post-years, starting in year three (a single percentage point pooled increase after turnaround). While pre-trend problems again limit causal warrants, this evidence suggests that takeover in Lawrence made the teaching force slightly more representative of the student population.

Holyoke. Unfortunately, the story was not especially positive for the second Massachusetts takeover which overall generated negative to neutral effects on student outcomes, at least in the first four years of reform (the period we examine here). We begin by displaying results for test-based outcomes in Figure 5. In math, the reforms did not appear to alter performance in the first year, but we observe negative effects in years two through four, resulting in a pooled impact of -0.22 SD over all four years. The pooled impacts were smaller but still negative in ELA (-0.06 SD) and science (-0.04 SD). There is some suggestive evidence that these results could be due to pre-trends on the test outcomes (all joint F-tests were statistically significant as reported in Table 6), but the findings were generally robust to the use of synthetic control methods as shown in Figure 5. When it came to the non-test outcomes, we find no strong evidence of impacts—positive or negative—on attendance, in-school suspensions, and grade progression. For out-of-school suspensions, Holyoke was on a very different pre-takeover trajectory than the comparison districts with very high rates of exclusionary discipline in the pre-reform era that plummeted the year prior to the reforms, making it difficult to draw conclusions about the impact of the reforms on out-of-school suspensions based on our event study methods (see Figure 6).

The negative impacts do not appear to be due to the changing composition of the Holyoke student population. Results are robust to the inclusion of student fixed effects, as we

show in Appendix Figure A7. Also, we observe no changes in the share of students identified as Black, Hispanic or low-income as a result of the reforms, as reported in Appendix Table A1. Therefore, it does not appear that these results are due to improvements in retention rates among disadvantaged students, for example.

In Table 7 we explore whether the Holyoke reform impacts varied depending on student demographic characteristics. In general, it did not appear that vulnerable subgroups benefited more from these reforms than other students. In fact, the negative effects were larger for Hispanic students, students from low-income homes, and students with a first language other than English. One bright spot is that the impacts were more neutral or even more positive for special education students than students not receiving special education services, especially when it came to test score outcomes.

Many of the Holyoke reforms targeted the district's youngest and oldest students, rather than those in the middle grades (which includes a large share of those students in tested grades who contribute to our estimates of the reform impacts on test scores). In Table 5, we explore whether the impacts of the reforms varied depending on student grade level. We do not find that it was only middle school students, for example, that were driving the negative results. In fact, for math, the negative impacts were no different for middle and high school students than for elementary students. In ELA, the effects were somewhat more negative for high schoolers than elementary students but no different for middle schoolers. For science, the negative impacts were concentrated among elementary school students. For attendance and in-school suspensions, middle schoolers saw more positive results than elementary schoolers.

We next turn to the impact of the Holyoke reforms on the characteristics of teachers working in the district. Results displayed in Figure 7 show that the reforms increased teacher turnover, driven both by teachers leaving teaching and/or the state—a five percentage point pooled increase—and transfering districts—an eleven percentage point pooled increase (as we show in Appendix Table A3). There was a particularly notable exodus of teachers out of the district in the first year after the reforms began although increased turnover occurred in all post-takeover years. However, it is not possible to rule out whether this was due to the reforms or due to pre-existing differences between Holyoke and comparison districts in teacher transfer rates during the pre-takeover period. Regardless, it appears the resulting teacher replacements led the Holyoke teaching force to become less experienced (a pooled decrease in experience of roughly two years), and these impacts increased over time. Finally, on a more positive note, the reforms increased the share of Black and Hispanic teachers (a pooled increase of two and one percentage points, respectively), making the teaching force somewhat more reflective of the student population than it was prior to the reforms.

SEZP. The Springfield Empowerment Zone Partnership reforms produced generally positive to neutral effects on student academic outcomes, both for the first and second cohorts. In Figure 8, we observe positive effects on all three test subjects by the second year of the reforms for Cohort 1. However, there are some signs that SEZP Cohort 1 was on a different pre-treatment trajectory than the comparison districts based on the joint F-tests reported in Table 3. The results based on synthetic control methods suggest positive effects for SEZP Cohort 1 by the fourth year of the reforms, but the positive effects in the earlier years are not always robust to this method, as we show in Figure 9. For Cohort 2, we also observe large positive impacts on all three test subjects, but again, there is evidence that this cohort was on a different trajectory than the comparison group in the pre-period. The positive effects on math and science are robust to

synthetic control methods but the ELA effects are not (see Figure 11). Student fixed effects estimates suggest that results were not driven by changes to the composition of the student population, as we show in Appendix Figure A7.

For Cohort 1 non-test outcomes, it is also difficult to differentiate treatment impacts from pre-existing differences in outcome trends. Our pooled difference-in-differences estimates in Table 10 suggest non-statistically significant, small, positive effects on attendance on the order of 0.58 days of school. These positive results appear robust to the use of synthetic control methods (see Figure 9). Similarly, we observe a one percentage point increase in the rate of grade progression as a result of the reforms, which is robust to synthetic control. Our pooled results suggest a small post-treatment reduction in in-school suspension rates, but the results are more neutral based on the synthetic comparison group. Unfortunately, the reforms appeared to increase out-of-school suspensions by a small 0.06 suspensions for Cohort 1 (pooling across all post-reform years), and this result persists even when relying on the synthetic comparison group. Cohort 2 students appeared to experience gains on all non-test outcomes we measured post-treatment, however, we again are unable to determine whether these effects are due to the reforms or pre-existing differences in SEZP Cohort 2 outcomes in the pre-treatment period. One exception is attendance which the reforms increased by 8.16 days, and this result is robust to the use of synthetic control methods (see Figure 12). While it was difficult to find patterns of variation for the SEZP results based on student demographic characteristics that were largely consistent across outcomes for Cohort 2, the results in general were more positive for Black students while the impacts were not as large for Hispanic students or for low-income students.

Pre-trends make it difficult to draw strong conclusions about the causal impacts of the SEZP reforms on teacher characteristics and workforce stability. However, some of the patterns appeared to be a bit different than those for the takeover districts. Specifically, the SEZP Cohort 1 reforms had no impact on the rate of teachers leaving teaching and/or the state. This result does not appear due to differences in the pre-reform trajectory of the SEZP Cohort 1 schools. Between-district transfers increased in both the year leading up to reform and the first year of the SEZP 1 reforms. Pooling all post-treatment years, we estimate a small increase in between-district transfers as a result of the reforms (see Appendix Table A3) followed by a return to pre-reform levels of district transfers in years two through four. Importantly, the SEZP schools had much higher rates of teachers transferring to other districts (17 percent) than the state as a whole or than any of the other reform contexts. It is difficult to say much about the impacts of the reforms themselves on the rest of the teacher characteristics we examine because SEZP appeared to be on a somewhat different trajectory than the comparison districts on most of these outcomes leading up to the turnaround.

Southbridge. Unfortunately, our results suggest that the most recent Massachusetts takeover in Southbridge generated negative to neutral impacts on student outcomes, at least through the first three years of reforms. On test outcomes shown in Table 13, we observe large negative effects on the order of -0.22 SD in math, -0.16 SD in ELA, and -0.29 SD in science when pooling all post-takeover years. Southbridge was on quite a different trajectory than comparison districts in the pre-takeover period on all outcomes, as we show visually in Figures 14 and 15. The formal joint F-tests of the pre-treatment coefficients reported in Table 12 also confirm these pre-trend differences were statistically significant. However, our findings on negative test score impacts are robust in all three subjects to the use of synthetic control methods

where we compare Southbridge students to a synthetic district that was on a nearly identical trajectory to Southbridge in the pre-takeover period.

For non-test outcomes, the Southbridge takeover appeared to result in reductions in attendance based on our event study estimates, but we observe violations of the parallel trends assumption, and these findings are not robust to the use of synthetic control methods. Our event study estimates suggest the takeover increased exposure to exclusionary discipline, increasing in-school suspensions and out-of-school suspensions by 0.03 and 0.04 suspensions, respectively. Despite the presence of pre-trend differences for Southbridge, the discipline results appear robust to the use of synthetic control methods, as shown in Figure 15. We find no impacts, positive or negative, on grade progression overall. In general, these results are not driven by changes to the composition of the student population as the results are robust to the inclusion of student fixed effects (see Appendix Figure A7).

Next, we turn to the question of whether the impacts of takeover in Southbridge varied for demographic subgroups of students. Results are displayed in Table 13. First, we find that the test score impacts were not as negative for Black students, and there were greater post-takeover reductions in suspensions for Black students. In contrast, the effects for nearly all outcomes were more negative for Hispanic students than non-Hispanic students. Negative results were also concentrated among students for whom English was not a first language. In Table 5 we examine whether effects varied by grade level. For math, we find that the negative effects were driven more by elementary school students than middle or high schoolers. Declines in ELA scores were driven more by elementary and high schoolers than middle schoolers, while declines in science achievement were driven by all three levels but were the largest among high school students. The increases in both in-school and out-of-school suspensions occurred mostly at the middle school level. That said, given the pre-trend issues, it is hard to draw strong conclusions about subgroups.

Finally, we examine how the Southbridge takeover impacted the characteristics of the teaching force. As illustrated in Figure 16, the reforms appeared to increase teacher turnover both in terms of teachers transferring districts and leaving teaching and/or the state. Specifically, as we show in Appendix Table A3, rates of between-district transfer increased by an average of sixteen percentage points in the post-period, though we do see differences in pre-period trends for this outcome. Pre-trends were less of a concern for rates of leaving teaching/the state, where turnaround appeared to increase departure rates by seven percentage points. This resulted in a somewhat less experienced teaching population (a pooled reduction of almost two years of experience) with a higher share of first year teachers (a pooled sixteen percentage point increase), although both estimates are subject to pre-trend issues. The reforms also appeared to increase the number of Hispanic teachers working in the district, especially in years two and three of the reforms.

Discussion

The previous literature on improving school districts labeled low-performing has largely focused on short-term impacts on test-based outcomes in initial reform contexts. Less is known about the sustainability of turnaround effects over time, accountability policy impacts on important non-test outcomes, and the replicability of district improvement success across contexts. This cross-case study of four state-initiated district improvement efforts in the context of Massachusetts begins to address these topics. Our medium-term results for the Lawrence takeover indicate that it is indeed possible for a system serving a high concentration of low-

income students of color to generate academic gains through state takeover and districtwide turnaround efforts. Lawrence leaders were able to generate positive effects on math and science performance, reduce student exposure to exclusionary discipline, and increase the grade progression rate. Our examination of these behavioral outcomes is especially important given the known possibility that accountability policy can sometimes result in improvements to high-stakes test outcomes to the detriment of other outcomes not part of the accountability system.

Although we find positive impacts of these reforms, our findings suggest that sustaining gains at the same level as the initial improvements can be challenging. In the case of Lawrence, this appeared especially true in the context of state- and district-level leadership turnover once the process of transitioning back to local control began. This finding suggests some parallels with what appears to have happened in the New Orleans public school system in the aftermath of the return from state takeover to a new version of local control where results were mostly sustained but still somewhat mixed depending on the outcomes in terms of whether improvements continued at the same levels (Childs et al., 2023). This suggests that leaders considering takeover or embarking on it should be planful about the state's exit strategy from the outset. Researchers should also devote more attention to learning about the transition out of state takeover so they can better inform policy in this area.

In terms of replicability, the results for the SEZP intervention suggest that it is indeed possible to replicate districtwide turnaround gains across contexts, as we observe suggestive evidence of positive impacts on most of the test-based and non-test outcomes we examine. That said, it is somewhat challenging to fully separate out policy impacts from pre-existing differences between the treated schools and the comparison group in this context. However, replication of the Lawrence results was not guaranteed and proved more challenging in the two other state takeover contexts—Holyoke and Southbridge—where the reforms appeared to negatively impact some (though not all) of the key student outcomes that we were able to study.

That said, it is important to note that we were only able to examine four years of post-takeover outcomes in Holyoke and three years in Southbridge. It is not impossible that reforms generated longer-term benefits that we are unable to observe. For example, leaders indicated during interviews we conducted that many of the core reforms in Southbridge were just getting underway during year three. Some of the earliest reforms involved establishing some basic systems—such as shifting from paper to digitized record keeping—that may not have paid off in terms of immediate student outcomes but may ultimately prove critical to setting a foundation for future improvement. It is possible that some of the preconditions necessary for replicating Lawrence's success may not have been present in all other targeted contexts. Furthermore, the measures we have available may not have captured important benefits of reforms pursued in these contexts. For example, Holyoke's focus on early childhood education and postsecondary/workforce preparation may not be well-evaluated, for example, by test score outcomes available only among the tested grades of 3-8 and 10. There may have been gains on outcomes beyond the scope of this study, such as access to pre-K and/or advanced course-taking, or high school graduation. For example, it does appear based on the publicly-available data that Holyoke experienced reductions in chronic absenteeism relative to the state as a whole.

Despite these limitations, it is worth considering whether there was variation in the policies pursued across the four contexts that could help explain differences in the impacts on student achievement. It is striking that leaders in the two districts with the more positive outcomes—Lawrence and SEZP—took a more similar policy approach than the other two districts. More specifically, they focused on increasing school-level autonomy paired with

central office supports, which has also appeared to be a potent combination in other contexts (e.g., Honig & Rainey, 2012; Jackson, 2023). Notably, leaders in these two districts also delayed high school reform. This may be a valuable recipe, especially when combined with other features known to be associated with improved academic outcomes in turnaround contexts, such as extended learning time (Schueler et al., 2021) and higher teacher pay in the context of a performance-based career ladder system (e.g., Dee & Wyckoff, 2015; Hanushek et al., 2023).

The impact of reforms on the characteristics of the teaching forces across each of the four contexts may also hint at the mechanisms through which the reforms operated. The two districts with less positive turnaround effects—Holyoke and Southbridge—experienced the greatest post-turnaround increases in teacher turnover, resulting in a less experienced workforce. In Lawrence, turnaround resulted in higher rates of teacher turnover and a less-experienced teaching population though the magnitudes of these effects were smaller than for Holyoke and Southbridge. It is possible that Lawrence was better able to make strategic staffing replacements than the other three districts due to its closer proximity to Boston. This indeed came up in our interviews with district and state leaders as a challenge for the districts with more negative outcomes, as was the related observation that it was challenging to convince charter operators to come work in the Western part of the state. Therefore, it may have been easier for Lawrence to rely on some teacher replacements to aid improvement (e.g., replacing less effective teachers with more effective ones), while for other districts the turnover may not have led to an overall more effective teaching force. In these communities it is possible that the more effective teachers had better opportunities outside of the turnaround districts and did not want to experience that takeover, causing them to leave at higher rates, undermining the reforms.

While we do not have definitive evidence of this possibility, the theory is consistent with previous research showing that teacher labor markets are highly localized (Engel & Cannata, 2015; Sanderson Edwards et al., 2022). Therefore, the ability for district leaders to rely on teacher replacements as an improvement strategy is likely dependent on the supply of potentially effective teachers in nearby districts or those nearby but not currently teaching. It is also possible that Lawrence leaders were simply better at identifying teachers who could or could not be successful in the turnaround context. Pham (2023) shows that teachers who leave turnaround schools in Tennessee context go on to be more effective in non-turnaround schools, suggesting context matters for teacher effectiveness. In other words, teachers do not simply sit on a binary continuum of effective to ineffective. Instead, some teachers are more effective in certain contexts than others and therefore administrators would be wise to attempt to identify teachers with the capacity to succeed specifically in school settings undergoing turnaround reform.

Another key takeaway from our research is that district and state leadership seemed to matter quite a bit when it came to the success of a districtwide improvement effort. This is suggested by the dips in Lawrence academic gains that coincided with the sudden death of the State Commissioner, the replacement of the original Superintendent/Receiver, and the transition to an appointed local board. There was also overlap in the groups involved in designing both the Lawrence and SEZP reforms more so than in the other contexts. Additionally, it was clear from our policy tracking work that frequent Superintendent churn made things particularly challenging in Southbridge. Furthermore, the state leadership's capacity for supporting multiple turnaround efforts could have contributed to challenges in Holyoke and Southbridge.

An underappreciated limitation of the accountability policy literature, is that it often focuses on the impact of the intervention itself rather than the impact of the accountability pressure (designed to incentivize better performance). For example, in the state takeover space,

although researchers have found that when takeover occurs it does not increase student achievement on average (Schueler & Bleiberg, 2021), it may be the case that state takeover laws improve achievement in low-performing systems via the threat of takeover in districts that never actually experience one. The SEZP experience provides suggestive evidence on this question as an example where the district, under the threat of takeover, avoided takeover by adopting a novel form of governance with greater state involvement than a typical district but greater local involvement than a typical takeover district. This arrangement benefited student achievement and is a model worth studying in more detail as it has now spread to ten different states covering over 29,000 students nationwide. Although more research is needed, this may be a path to district improvement that avoids some of the more contentious aspects of state takeover. This is particularly relevant to the state's ongoing consideration of laws that would alter the Achievement Gap Act's state takeover provisions.

One limitation of our work is that Massachusetts is a relatively high-performing state (Papay et al., 2020) and therefore the lowest-performing districts in Massachusetts are not the lowest-achieving in the country. As a result, our findings may not generalize to other contexts, particularly those within the very lowest-performing states in the country. This is especially true because it appears that the impact of turnaround reforms has historically been larger in higher-performing contexts than those contexts that are relatively higher-performing in the national performance distribution (Schueler, 2023).

Another sobering point is that both the Lawrence and SEZP reforms narrowed gaps in achievement between themselves and the statewide average but unfortunately did not close these gaps entirely. Lawrence caught up on some outcomes with other majority low-income districts but, by the end of the seven years of reform, was still performing well behind the average for the state as a whole. SEZP made notable gains but remained achieving at levels below that of other majority low-income districts in the state as of the last pre-pandemic year. This research provides valuable guidance for making non-trivial and important improvements of a magnitude not often observed in school systems serving high concentrations of low-income children of color. However, there remains an urgent need to identify strategies capable of more fully addressing opportunity gaps and ultimately eliminating educational inequality.

Despite this need, the magnitude of the positive impacts in Lawrence and SEZP were still large and noteworthy. In math achievement, the effects were equivalent to roughly one-quarter of the overall average statewide difference in achievement between low-income and non-low-income students. These impacts were comparable in size to the impact of efforts to implement the practices of high-performing charter schools into traditional public schools in Houston (Fryer, 2014) and to grandfather traditional public school students into high-performing charter schools in New Orleans (Abdulkadiroglu et al., 2014). Therefore, although these districts still have ample room to grow, leaders did generate rare and remarkable improvements in student academic outcomes, providing lessons for leaders seeking to do the same in their own communities.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495-510.

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391-425.

Backes, B. & Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, 68, 89-103.

Bailey, M. & Dynarski, S. (2011). "Inequality in postsecondary education" in Greg Duncan and Richard Murnane eds., *Whither opportunity: Rising inequality, schools, and children's life chances.* New York, NY: Russell Safe Foundation.

Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). *Facing the Challenges of Whole-School Reform: New American Schools After a Decade*. Rand Corporation.

Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, *95*(3), 983-1001.

Blazar, D. & Schueler, B. (2022). Why Do School Districts Matter? An Interdisciplinary Framework and Empirical Review. Annenberg Working Paper No. 22-581. https://www.edworkingpapers.com/ai22-581

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive School Reform and Achievement: A Meta-Analysis. *Review of Educational Research*, *73*(2), 125–230.

Carter, P.L., & Reardon, S.F. (2014). Inequality Matters: Framing a Strategic Inequality Research Agenda. *The William T. Grant Foundation*.

Childs, Carroll, Harris, Tracy, Shields, Jabbar, Marsh & Ogden (2023). Back to the Future? From State Takeover to the Nation's First All-Charter, District-Governed School System in New Orleans. Working Paper.

Chin, M., Kane, T. J., Kozakowski, W., Schueler, B. E., & Staiger, D. O. (2019). School District Reform in Newark: Within- and Between-School Changes in Achievement Growth. *ILR Review*, *72*(2), 323–354.

Chingos, M. M., & Grover J. Whitehurst & Michael R. Gallaher, (2015). School Districts and Student Achievement. Education Finance and Policy, MIT Press, vol. 10(3), p. 378-398, July.

Commonwealth of Massachusetts (2010). Chapter 12: An Act Relative to the Achievement Gap. https://malegislature.gov/Laws/SessionLaws/Acts/2010/Chapter12

Dee, T. & Wyckoff, J. (2015). Incentives, Selection, and Teacher Performance: Evidence from from IMPACT. *Journal of Policy Analysis and Management*, 34(2), 267-297.

Deming, D., Cohodes, S., Jennings, J. & Jencks, C. (2016). School Accountability, Postsecondary Attainment, and Earnings. *Review of Economics and Statistics*, 98(5), 848-862.

Desimone, L. (2002). How Can Comprehensive School Reform Models Be Successfully Implemented? *Review of Educational Research*, *72*(3), 433–479.

Duncan, G. J., & Murnane, R. J. (2011). *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation.

Empower Schools (2023). Our approach: National in scope, locally driven. https://www.empowerschools.org/our-approach/#where-we-work

Engel, M. & Cannata, M. (2015). Localism and Teacher Labor Markets: How Geography and Decision Making May Contribute to Inequality. *Peabody Journal of Education*, 90, 84-92.

Eren, O., Figlio, D., Mocan, N. & Ozturk, O. (2023). School accountability, long-run criminal activity, and self-sufficiency. NBER Working Paper No. 31556.

Fried, S. (2020). State takeover: Managing emotions, policy implementation, and the support/sanction duality in the Holyoke Public Schools Receivership. *Harvard Educational Review*, 90(1), 75-101. https://meridian.allenpress.com/her/article/90/1/75/432554/State-Takeover-Managing-Emotions-Policy

Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). *State Takeover, School Restructuring, Private Management, and Student Achievement in Philadelphia*. RAND Corporation. https://www.rand.org/pubs/monographs/MG533.html

Gross, B., Booker, T. K., & Goldhaber, D. (2009). Boosting Student Achievement: The Effect of Comprehensive School Reform on Student Achievement. *Educational Evaluation and Policy Analysis*, *31*(2), 111–126.

Hanushek, E. A., Light, J. D., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2019). *Long-run Trends in the U.S. SES-Achievement Gap*. National Bureau of Economic Research.

Hanushek, E., Luo, J., Morgan, A., Nguyen, M., Ost, B., Rivkin, S. & Shakeel, A. (2023). The effects of comprehensive educator evaluation and pay reform on achievement. Annenberg Institute Working Paprt No. 23-768.

Hashim, S. A., Kane, T. J., Kelley-Kemple, T., Laski, M. E., & Staiger, D. O. (2020). *Have Income-Based Achievement Gaps Widened or Narrowed?* National Bureau of Economic Research.

Harbatkin, E., McIlwain, A. & Strunk, K. (2022). School Turnaround in a Pandemic: An Examination of the Outsized Implications of COVID-19 on Low-Performing Schools and Their Communities. Education Policy Innovation Collaborative.

Harris, D., & Larsen, M. (2022). Taken by Storm: The Effects of Hurricane Katrina on Medium-term Student Outcomes in New Orleans. Journal of Human Resources, 58(5).

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Henry, G., Pham, L., Kho, A. & Zimmer, R. (2020). Peeking Into the Black Box of School Turnaround: A Formal Test of Mediators and Suppressors. *Educational Evaluation and Policy Analysis*, 42(2), 232-256.

Hernández, M. (2019). Is there no excuse? The effects of the New Orleans school reforms on school discipline. *Education Research Alliance for New Orleans*.

Honig, M. & Rainey, L. (2015). Autonomy and School Improvement: What Do We Know and Where Do We Go From Here? *Educational Policy*, 26(3), 465-495.

Jackson, C. K. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *The Journal of Political Economy*, *126*(5), 2072–2107.

Jackson, K. (2023). When does school autonomy improve student outcomes? Brown University's Annenberg Institute EdWorkingPaper 23-808.

Jochim & Opalka (2017). The "City of Firsts" Charts a New Path on Turnaround. Center for Reinventing Education.

Johnson, S. M., Marietta, G., Higgins, M. C., Mapp, K. L., & Grossman, A. (2015). *Achieving Coherence in District Improvement: Managing the Relationship Between the Central Office and Schools*. Harvard Education Press.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement. *Educational Researcher*, *49*(8), 549–565. https://doi.org/10.3102/0013189X20965918

Lyon, M., Bleiberg, J. & Schueler, B. (2022). Does Centralization Promote Fiscal Health? The Effect of State Takeover of Local School Districts on Education Finance. Annenberg Working Paper No. 22-689.

McClelland, R., & Gault, S. (2017). The synthetic control method as a tool to understand state policy. *Washington, DC: Urban-Brookings Tax Policy Center*.

Meyers, C. V., Brandt, W. C., & VanGronigen, B. A. (2022). State ESSA Plans and Comprehensive Support and Improvement (CSI) Status. *Educational Policy*, 08959048221087209. https://doi.org/10.1177/08959048221087209

Moore Johnson, S. (2017). Investing in Teachers: The Lawrence Public Schools Respond to State Receivership. Cambridge, MA: Harvard Education Press.

Moore Johnson, S. (2021). Leaders Change, Policies Evolve. The Lawrence Public Schools Respond to State Receivership, Act II. Cambridge, MA: Harvard Education Press.

Papay, J. Mantil, A., Murnane, R., An, L., Donohue, K. & McDonough, A. (2020). Lifting all boats? Accomplishments and challenges from 20 years of education reform in Massachusetts. https://annenberg.brown.edu/edopportunity/liftingallboats

Papay, & Hannon. (2018). The effects of school turnaround strategies in Massachusetts. *Brown University Working Paper*. https://pdfs.semanticscholar.org/5f31/90a1d41b22e8ae4cbae4058194e52dad75ad.pdf

Peurach, D. J., & Neumerski, C. M. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, *16*(4), 379–420.

Pham, L. D., Henry, G. T., Kho, A., & Zimmer, R. (2020). Sustainability and maturation of school turnaround: A multiyear evaluation of Tennessee's Achievement School District and local Innovation Zones. *AERA Open*, *6*(2), 233285842092284.

Pham, L. (2023). Teachers Are Not Lemons: An Examination of Spillover Effects When Teachers Transfer Away From Turnaround Schools. *Educational Researcher*.

Redding, C., & Nguyen, T. D. (2020). The Relationship Between School Turnaround and Student Outcomes: A Meta-Analysis. *Educational Evaluation and Policy Analysis*, *42*(4), 493–519.

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor. In D. B. Grusky & J. Hill (Eds.), *Inequality in the 21st Century* (1st ed., pp. 177–189). Routledge. https://doi.org/10.4324/9780429499821-33

Sanderson Edwards, D., Kraft, M., Christian, A. & Candelaria, C. (2022). Teacher Shortages: A Unifying Framework for Understanding and Predicting Vacancies. EdWorkingPaper No. 22-684.

Schnurer, E. (2017). The Springfield Empowerment Zone Partnership. Progressive Policy Institute.

Schueler, B. E., Asher, C. A., Larned, K. E., Mehrotra, S., & Pollard, C. (2021). Improving Low-

Performing Schools: A Meta-Analysis of Impact Evaluation Studies. *American Educational Research Journal*, 00028312211060855.

Schueler, B. E., & Bleiberg, J. (2022). Evaluating Education Governance: Does State Takeover of School Districts Affect Student Achievement? *Journal of Policy Analysis and Management*, *41*(1), 162–192.

Schueler, B. E., & Goodman, J. S. (2017). Can states take over and turn around school districts? Evidence from Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*. https://journals.sagepub.com/doi/abs/10.3102/0162373716685824

Schueler, B. (2019). A Third way: The politics of school district takeover and turnaround in Lawrence, Massachusetts. *Educational Administration Quarterly*, 55(1), 116-153.

Schueler, B. (2018). Making the most of school vacation: A field experiment of small group math instruction. *Education Finance and Policy*, 15(2), 310-331.

Schueler, B. (2023). Where Are the Lowest-Scoring Schools in the United States? National Versus Within-State School Rankings and the Generalizability of School Improvement Research. Working Paper.

Schwartz, N., Kang, H., Loeb, S., Grissom, J., Bartanen, B., Cheatham, J., Chi, O., Donaldson, M., Lemos, R.F., Mellon, G., Moffitt, S., Nurshatayeva, A., Owens, J., Pinker, E., White, R., & Zimmerman, S. (2023). Studying the Superintendency: A Call for Research. Annenberg Institute at Brown.

Shores, K. A., Candelaria, C. A., & Kabourek, S. E. (2022). Spending more on the poor? A comprehensive summary of state-specific responses to school finance reforms from 1990–2014. *Education Finance and Policy*, 1-28.

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

Wong, K. K., & Shen, F. X. (2001). Does school district takeover work? Assessing the effectiveness of city and state takeover as a school reform strategy. Paper presented at the annual meeting of the American Political Science Association, San Francisco, CA.

Wong, K. K., and Shen, F.X. (2002). City and State Takeover as a School Reform Strategy, ERIC Clearinghouse on Urban Education,

Zimmer, R., Kho, A., Henry, G., & Viano, S. (2015). Evaluation of the Effect of Tennessee's Acheievment School District on Student Test Scores. *ASD Report*.

2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
Leadership of the State Department of Elementary and Secondary Education						
Commissioner 1					Acting/Interim Commissioner	Commissioner 2
Lawrence Receivership						
Receiver-Superintendent 1						Superintendent 2
					Receiver Board 1	
			Holyoke Receivership			
			Receiver-Superintendent 1			
			Springfield Empowerment Zone Partnership			
			Board of Directors Chair 1 / Superintendent 1			
				Southbridge Receivership		
				Receiver 1	Interim Receiver 2 / Receiver 3	Receiver 3

Figure 1a. Timeline of State-led District Turnaround in Massachusetts.

	Lawrence	Holyoke	SEZP	Southbridge
Interventions				
First Year	2012-13	2015-16	2015-16	2016-17
Leadership	Same Receiver until last year, coincided w/ gas explosions	Same Receiver over whole period	Same Board of Directors Chair & Superintendent over whole period	Three Receivers over three years
Autonomy & Accountability	Increased (differentiated) school autonomy Small set of non-charters managed by operators	Somewhat increased school autonomy Small set of non-charters managed by operators	Increased (differentiated) school autonomy Small set of non-charters managed by operators	
Extended Time	Extended year/day Vacation Academies HS high-dose tutoring	Extended year/day Vacation Academies	Extended year/day Vacation Academies High-dose tutoring	Extended year/day for elementary
Human Capital	Principal and teacher replacements Performance-based career ladder Raised pay Teacher leadership teams	Principal and teacher replacements Performance-based career ladder Raised pay New principal supervisor roles School leader capacity building	Principal and teacher replacements Performance-based career ladder Raised pay Teacher leadership teams School leader capacity building	Principal and teacher replacements Performance-based career ladder Raised pay Increased PD time Sheltered English immersion teacher training
Data Use	Increased data use	Increased data use	Increased data use	Increased data use Shift from paper to digital record keeping
Family Engagement	Family resource center Partners connecting families with employers	Expanded pre-K Parent advisory council New translation/ interpretation services Expanded home visiting Family resource center		Family liasons at all schools New translation/ interpretation services Hired director of communications
Instruction / Curriculum	Built out enrichment offerings New standards-aligned curriculum (midway)	Built out enrichment offerings New standards-aligned curriculum (midway) Dual language program HS advanced coursework / dual enrollment Technology-enhanced personalized learning	New standards-aligned curriculum New dual language programs New early college program	New standards-aligned curriculum Dual language program
Climate				Positive Behavioral Interventions & Supports districtwide
High School Redesign	New school for at-risk students High school redesign (midway)	High school redesign New alternative HS pathways	High school added to the zone in 2017-18	Expanded credit recovery offerings New alternative therapeutic day middle/high school
Results				
Math	Positive	Negative	Positive	Negative
ELA	Neutral	Negative	Positive/Neutral	Negative
Science	Positive	Negative	Positive	Negative
Attendance	Positive	Neutral	Positive	Negative/Neutral
In-School Suspensions	Positive	Neutral	Positive/Neutral	Negative
Out-of-School Suspensions	Positive/Neutral	Neutral	Negative	Negative
Grade Progression	Positive/Neutral	Neutral	Positive	Neutral

Figure 1b. Summarizing and Comparing Interventions and Results Across Contexts.

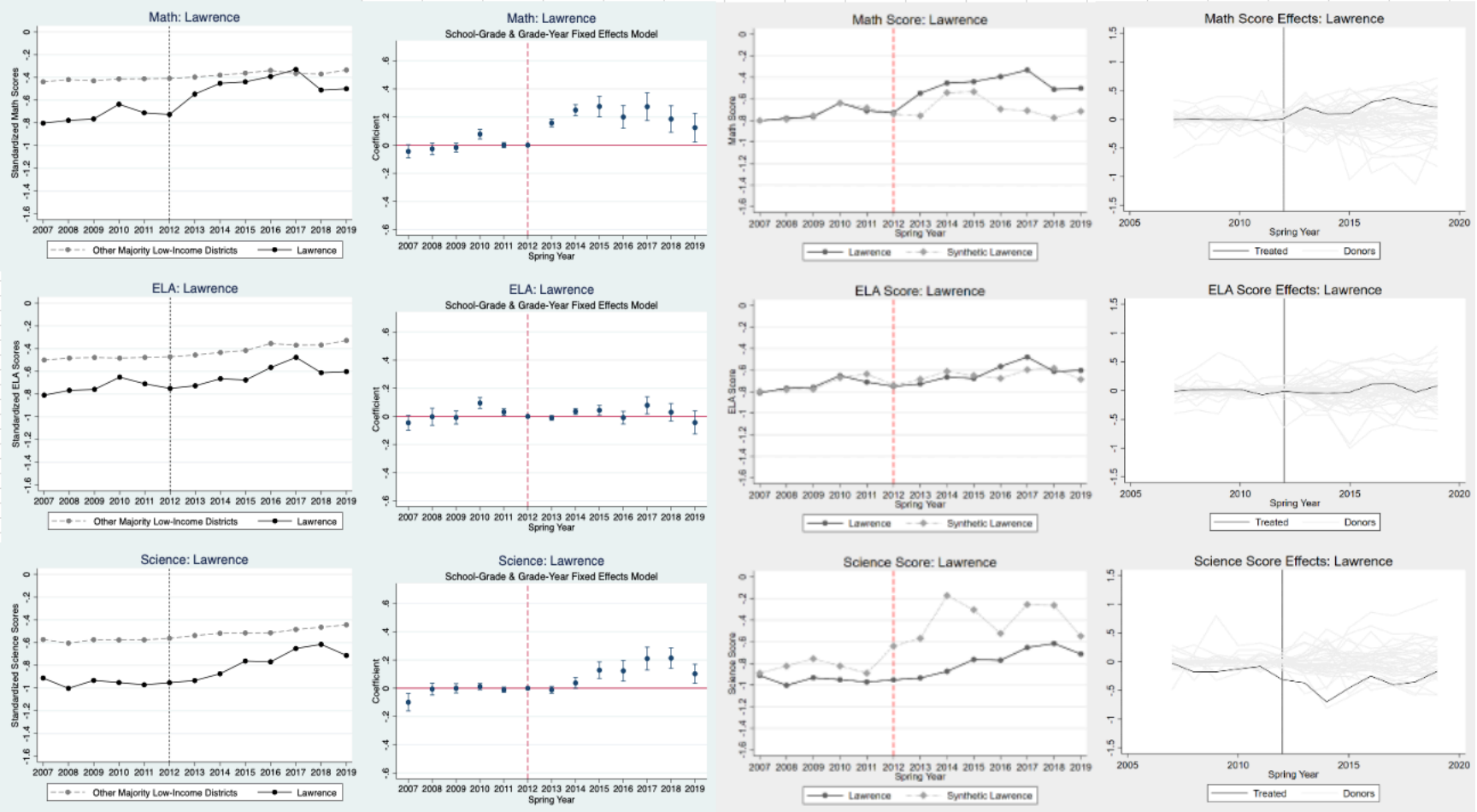

Figure 2. Lawrence test-based outcome trends, event study estimates, synthetic control estimates, and placebo tests.

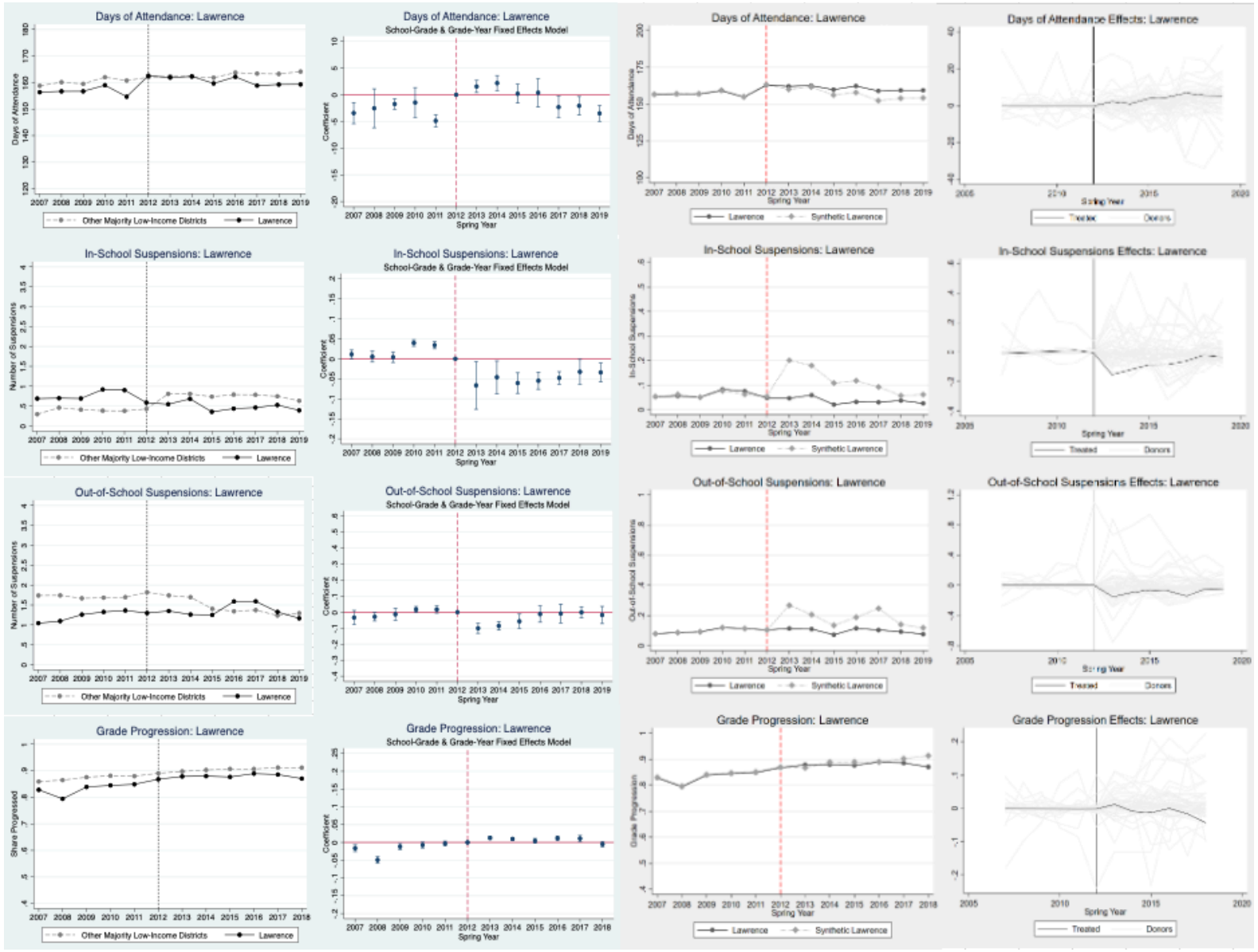

Figure 3. Lawrence non-test outcome trends, event study estimates, synthetic control estimates, and placebo tests.

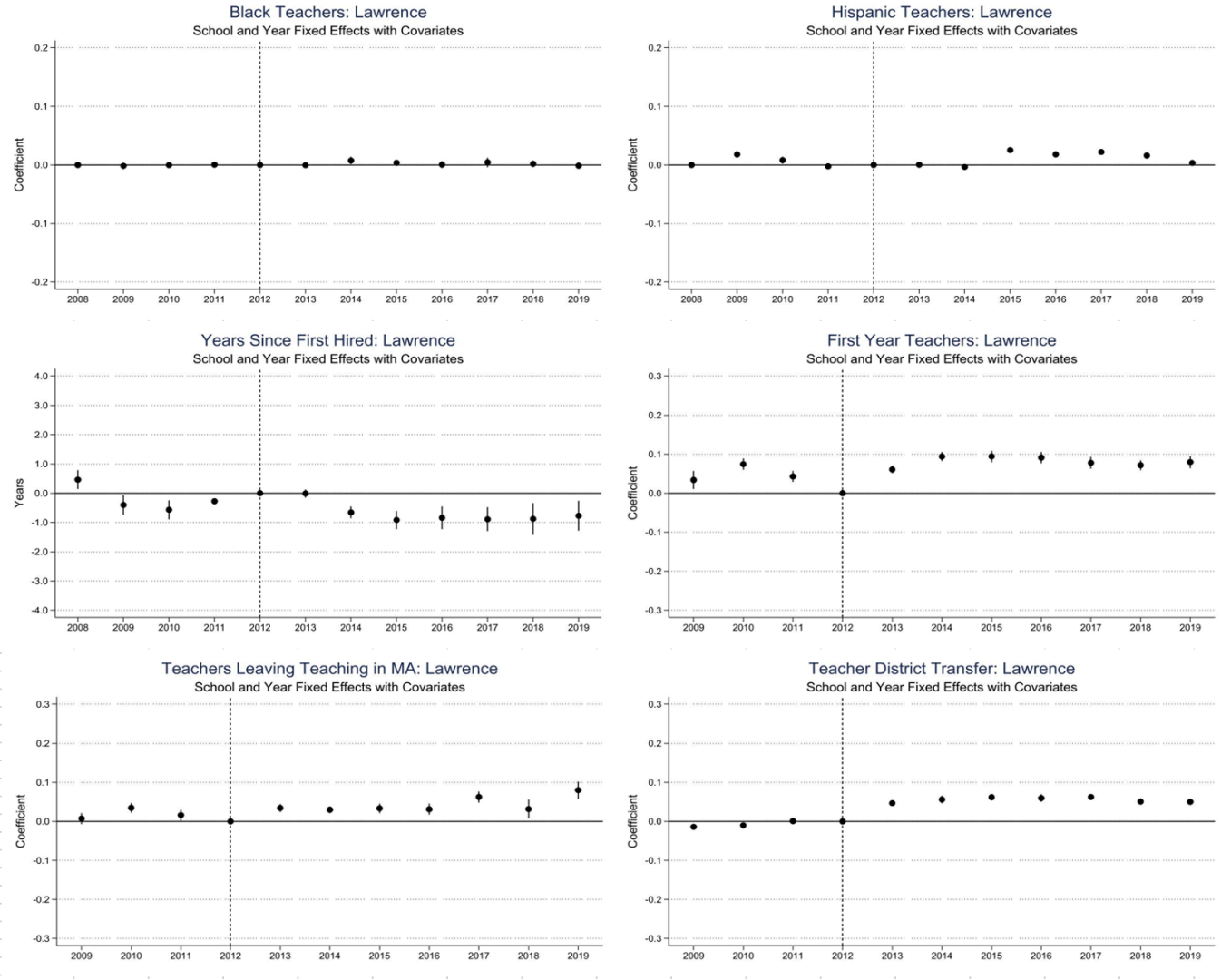

Figure 4. Lawrence reform impacts on teacher characteristics, event study estimates.
Note: The 2008 year is censored for the first year teacher and turnover variables. The turnover variables are backward-looking (e.g., turnover rates for 2013 reflect whether teachers who were present in 2011-12 were also present in 2012-13).

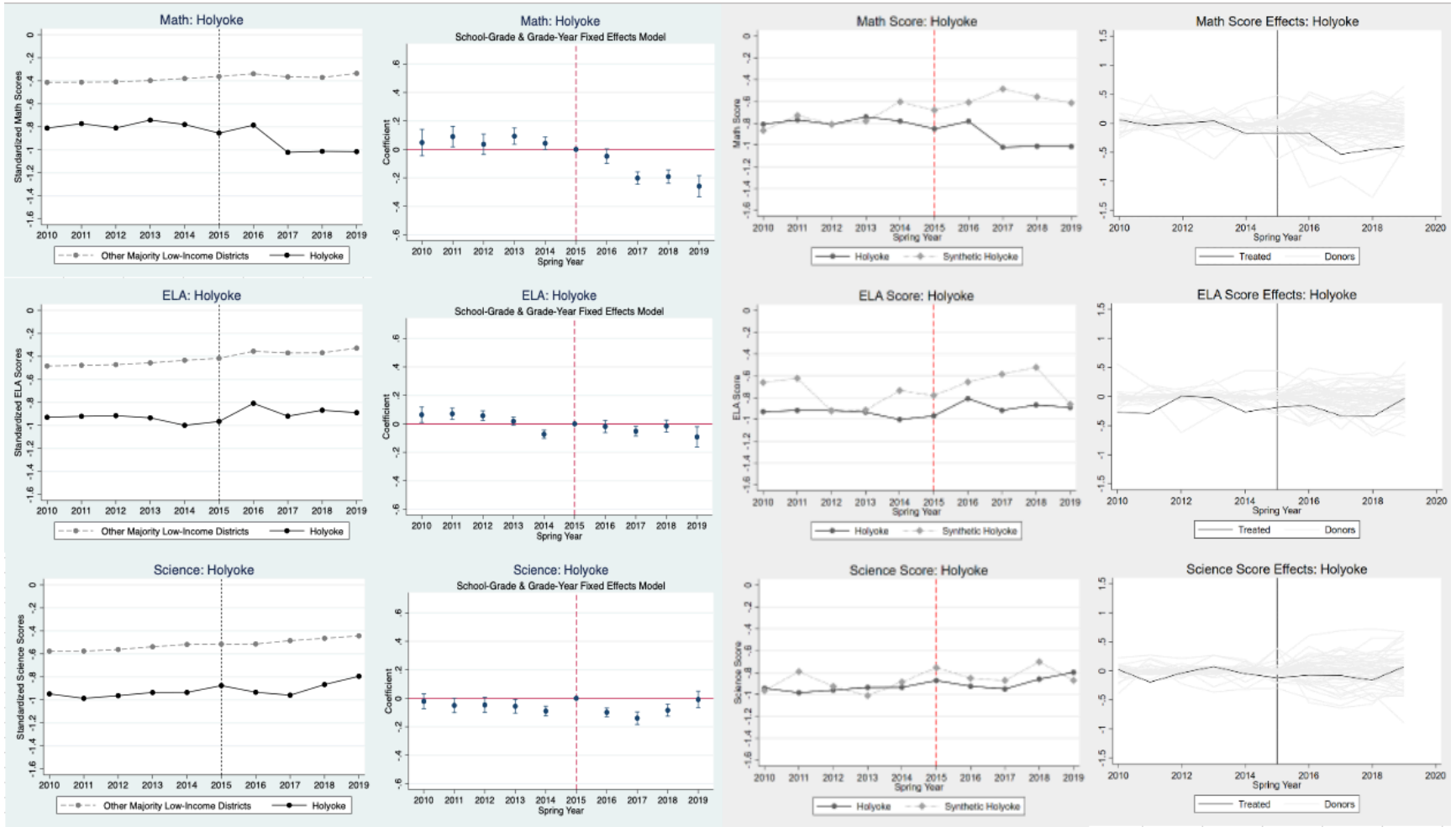

Figure 5. Holyoke test-based outcome trends, event study estimates, synthetic control estimates, and placebo tests.

Figure 6. Holyoke non-test outcome trends, event study estimates, synthetic control estimates, and placebo tests.

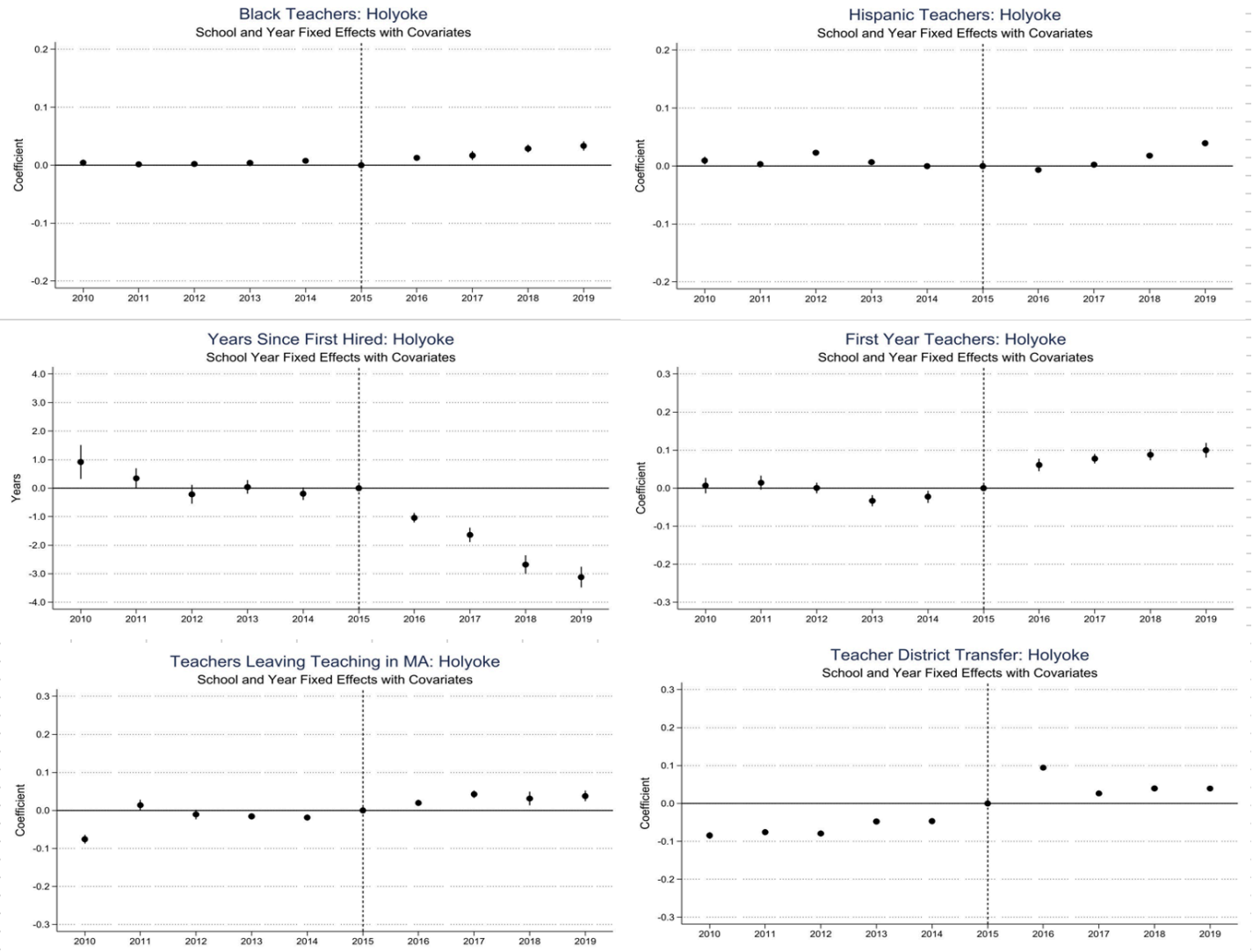

Figure 7. Holyoke reform impacts on teacher characteristics, event study estimates.
Note: The turnover variables are backward-looking (e.g., turnover rates for 2013 reflect whether teachers who were present in 2011-12 were also present in 2012-13).

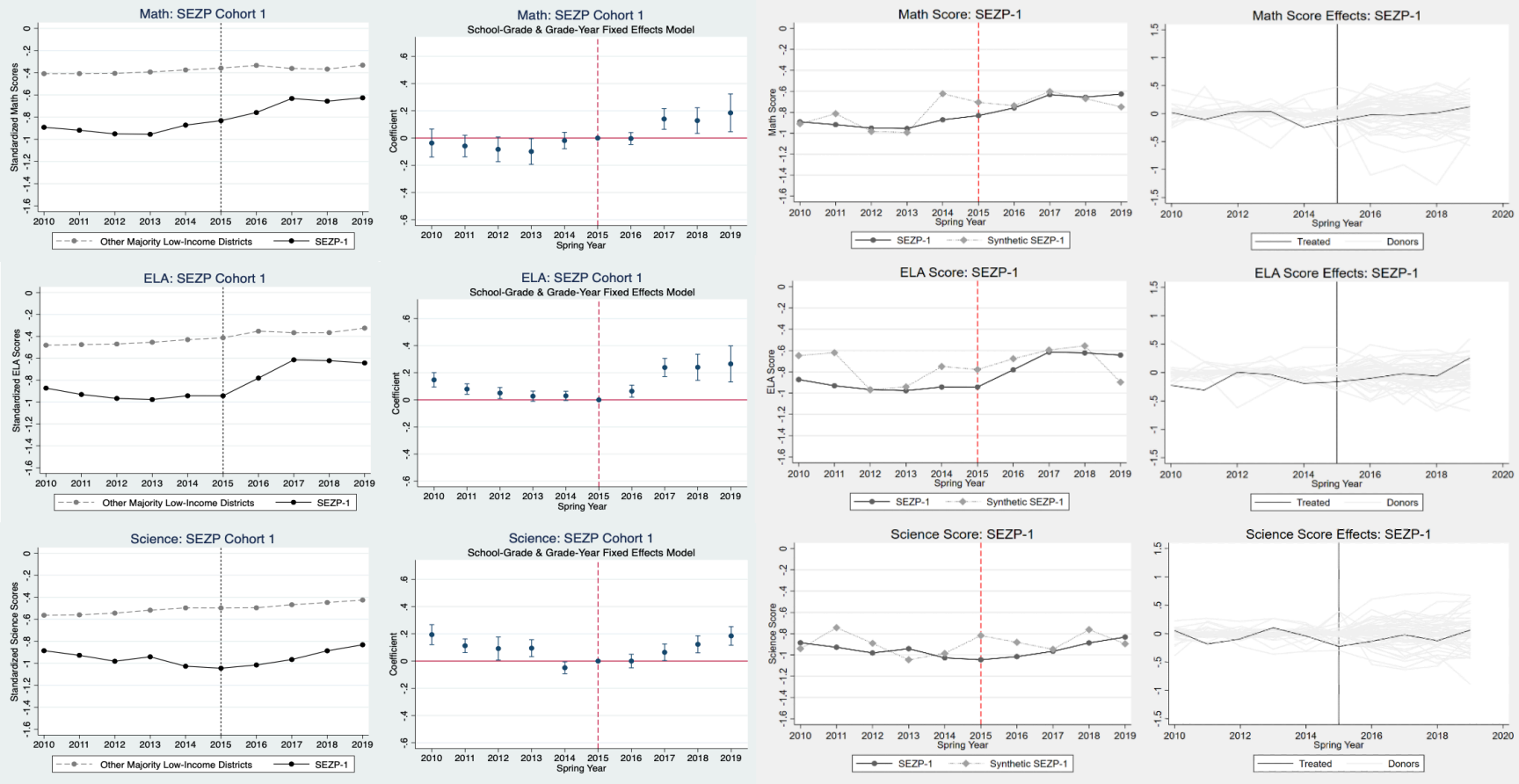

Figure 8. SEZP Cohort 1 test-based outcome trends, event study estimates, synthetic control estimates, and placebo tests.

Figure 9. SEZP Cohort 1 non-test outcome trends, event study estimates, synthetic control estimates, and placebo tests.

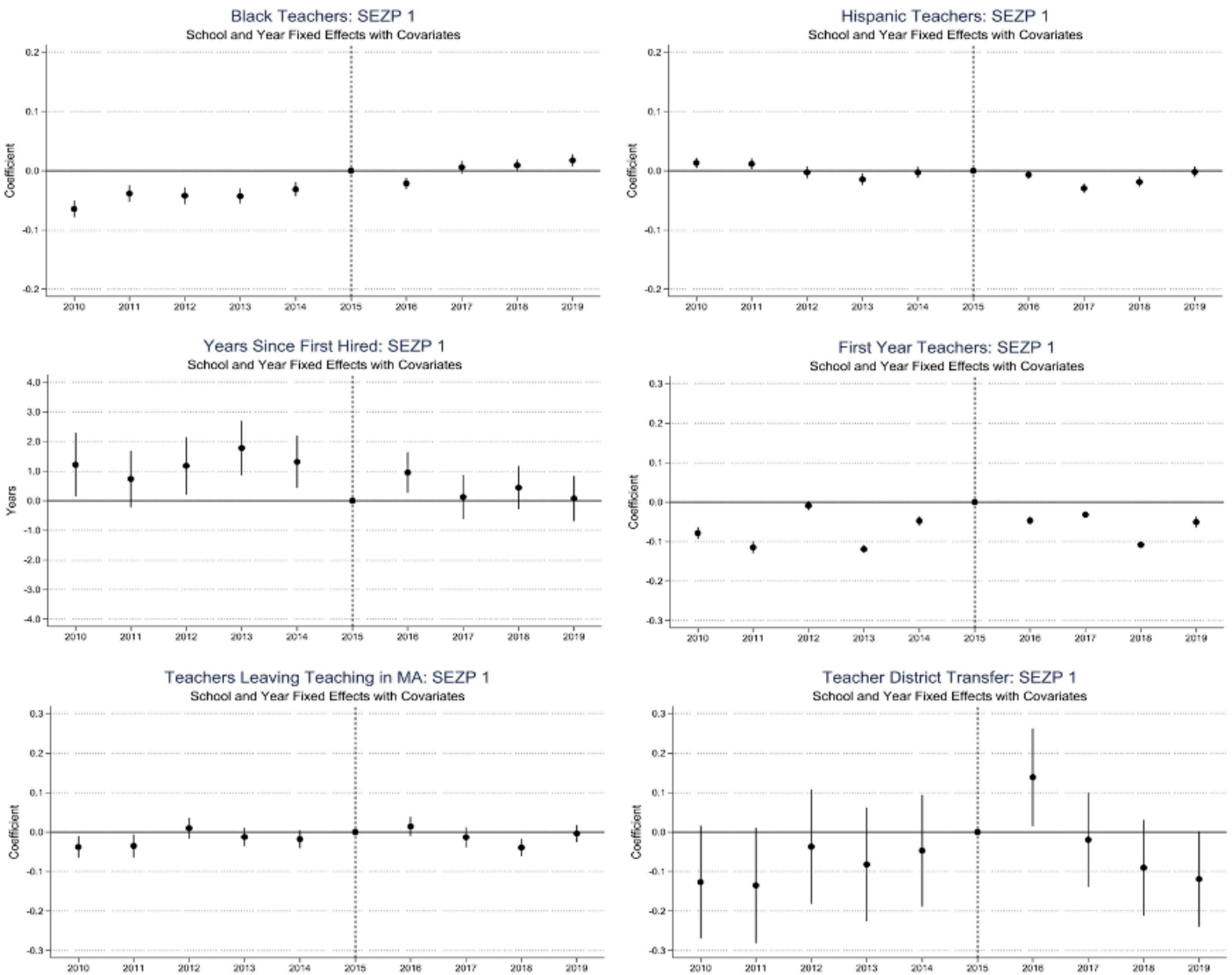

Figure 10. SEZP Cohort 1 reform impacts on teacher characteristics, event study estimates.
Note: The turnover variables are backward-looking (e.g., turnover rates for 2013 reflect whether teachers who were present in 2011-12 were also present in 2012-13).

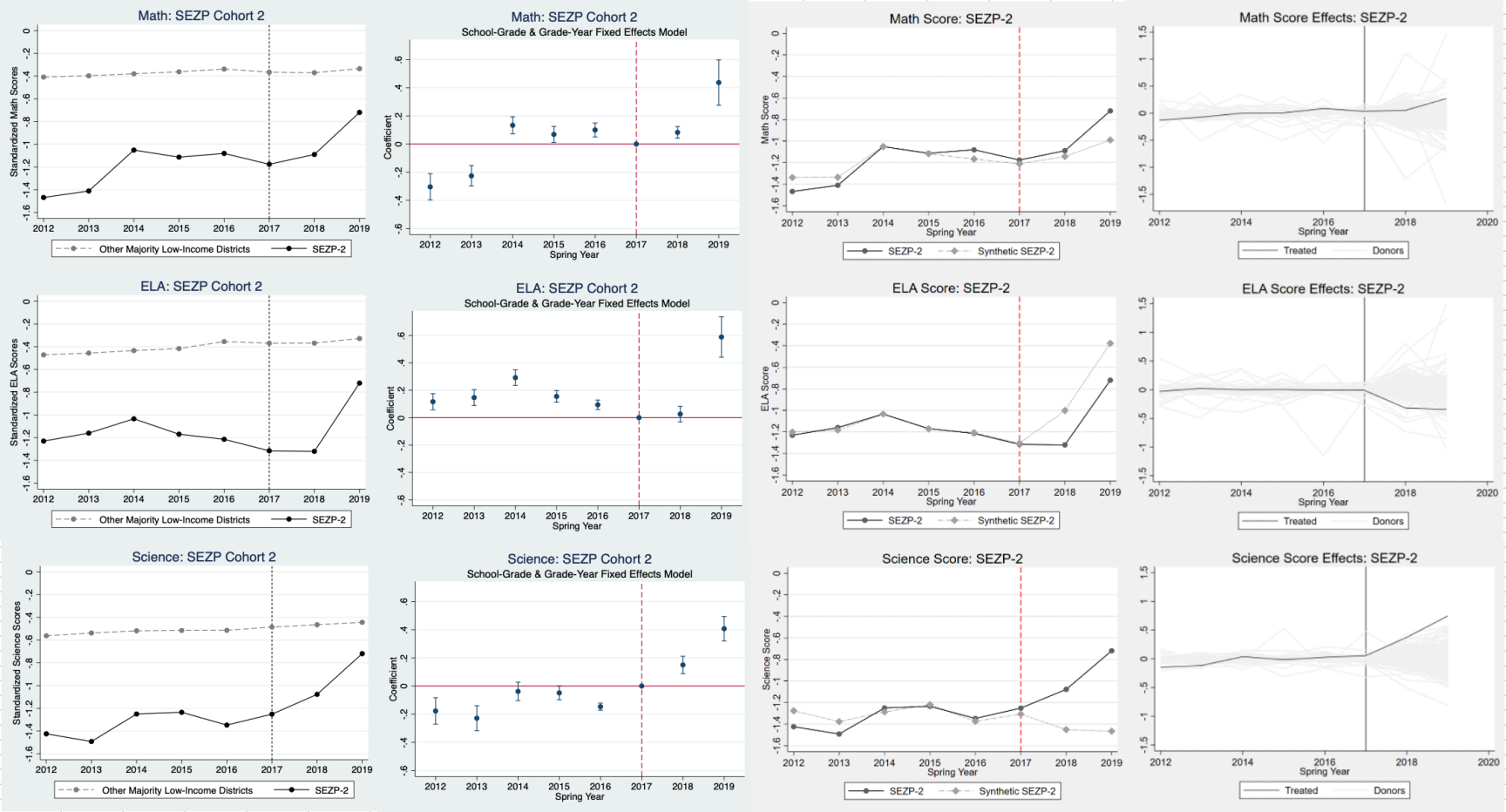

Figure 11. SEZP Cohort 2 test-based outcome trends, event study estimates, synthetic control estimates, and placebo tests.

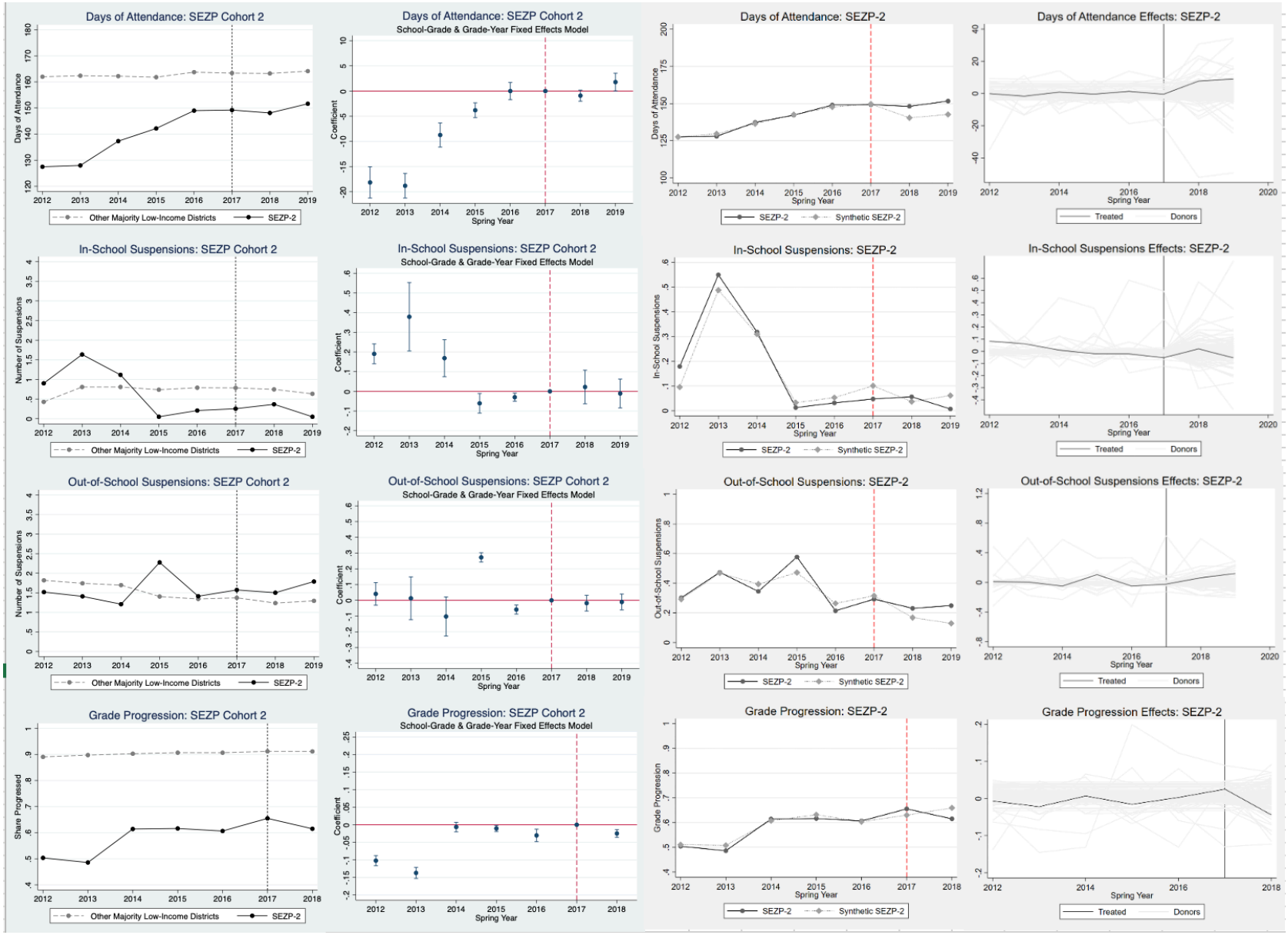

Figure 12. SEZP Cohort 2 non-test outcome trends, event study estimates, synthetic control estimates, and placebo tests.

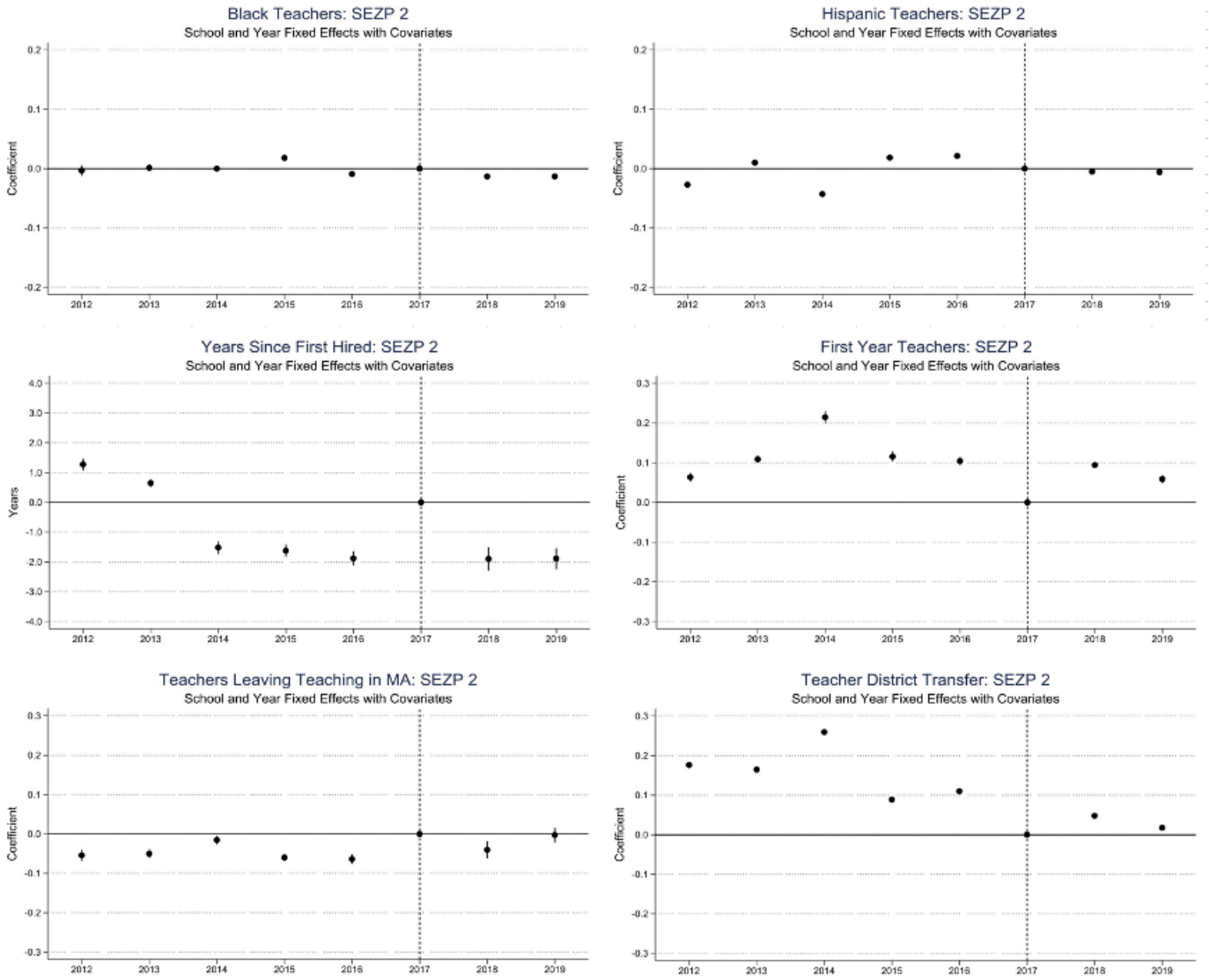

Figure 13. SEZP Cohort 2 reform impacts on teacher characteristics, event study estimates.
Note: The turnover variables are forward-looking (e.g., turnover rates for 2013 reflect whether teachers who were present in 2011-12 were also present in 2012-13).

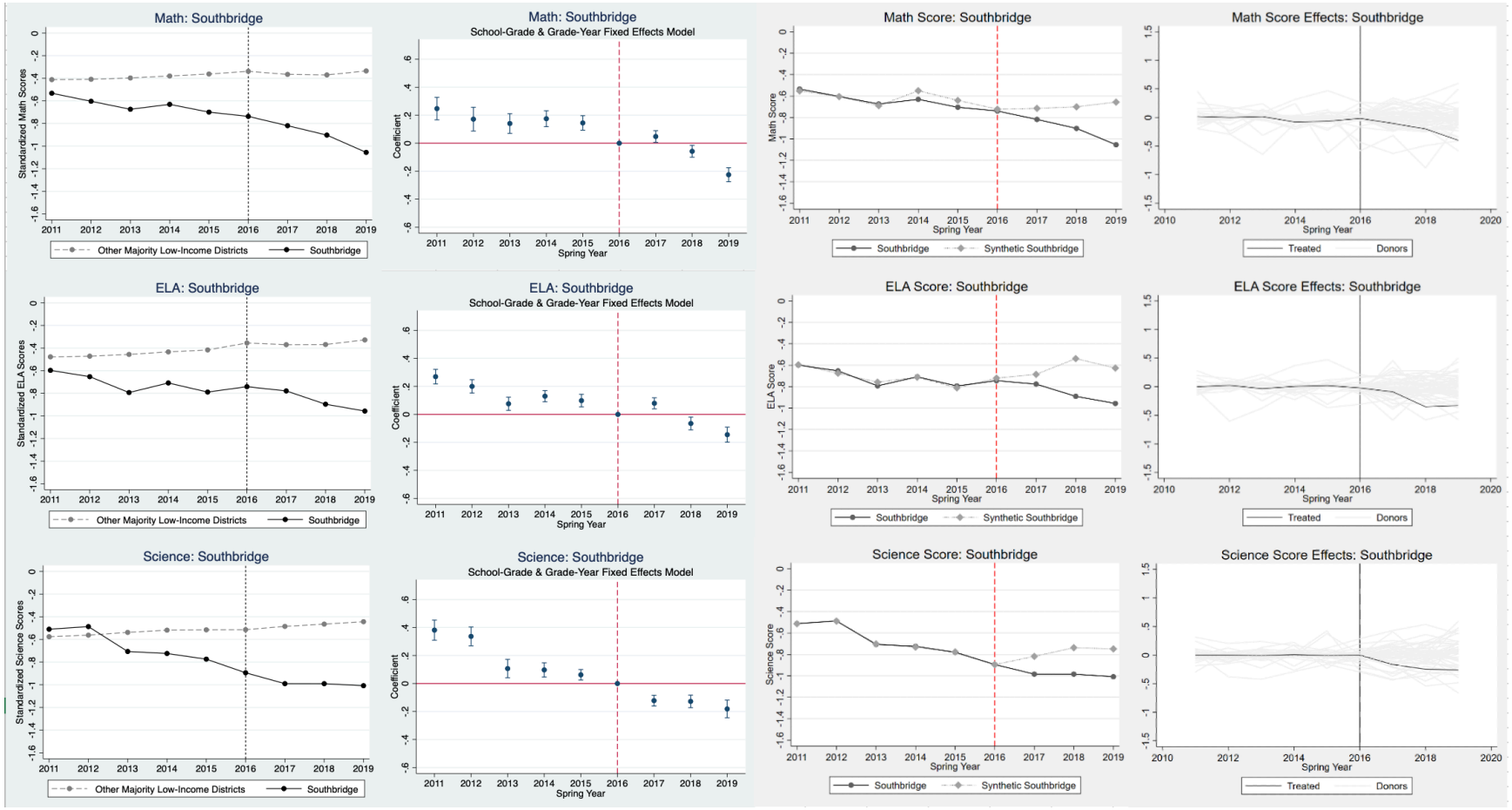

Figure 14. Southbridge test-based outcome trends, event study estimates, synthetic control estimates, and placebo tests.

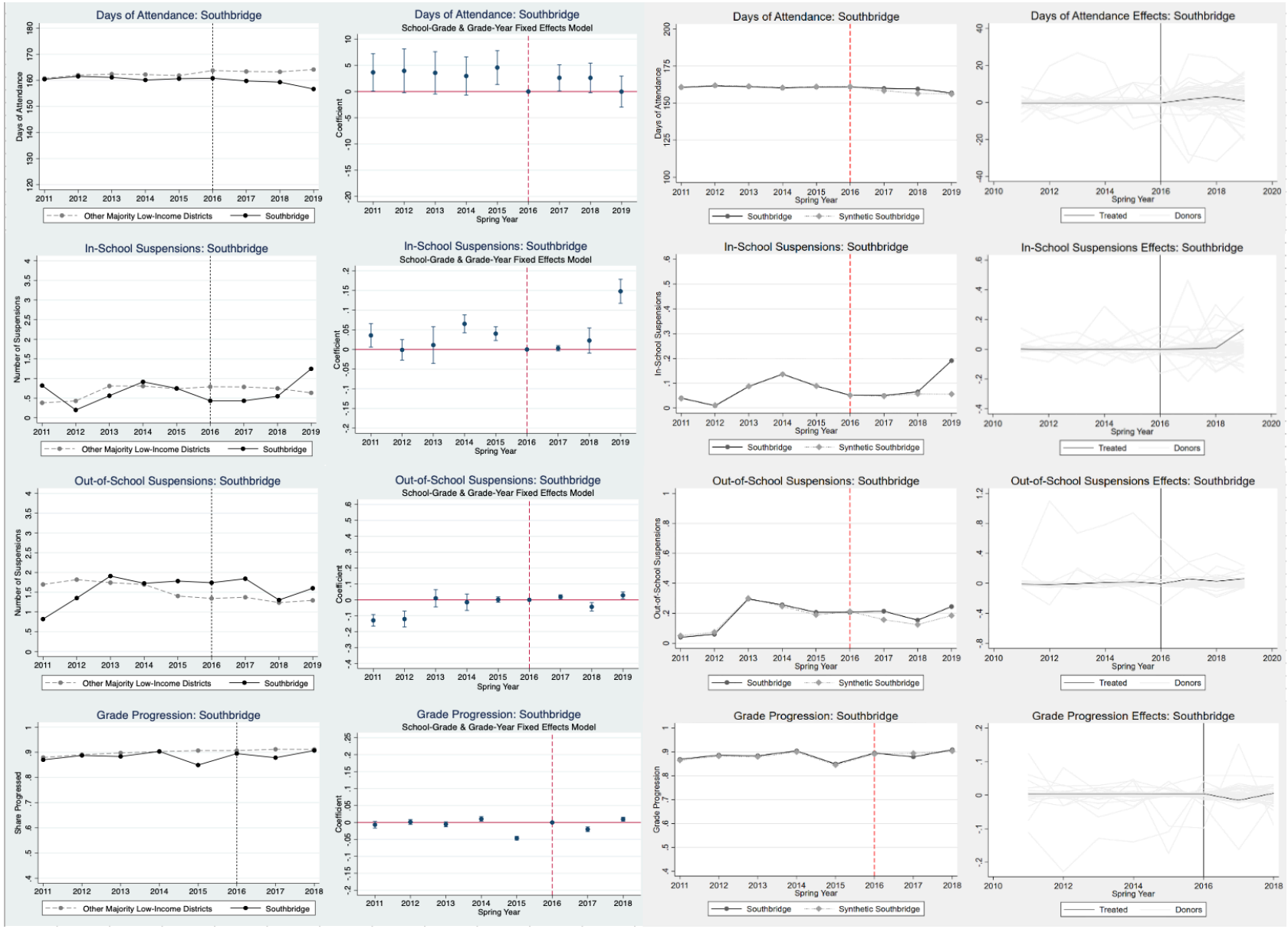

Figure 15. Southbridge non-test outcome trends, event study estimates, synthetic control estimates, and placebo tests.

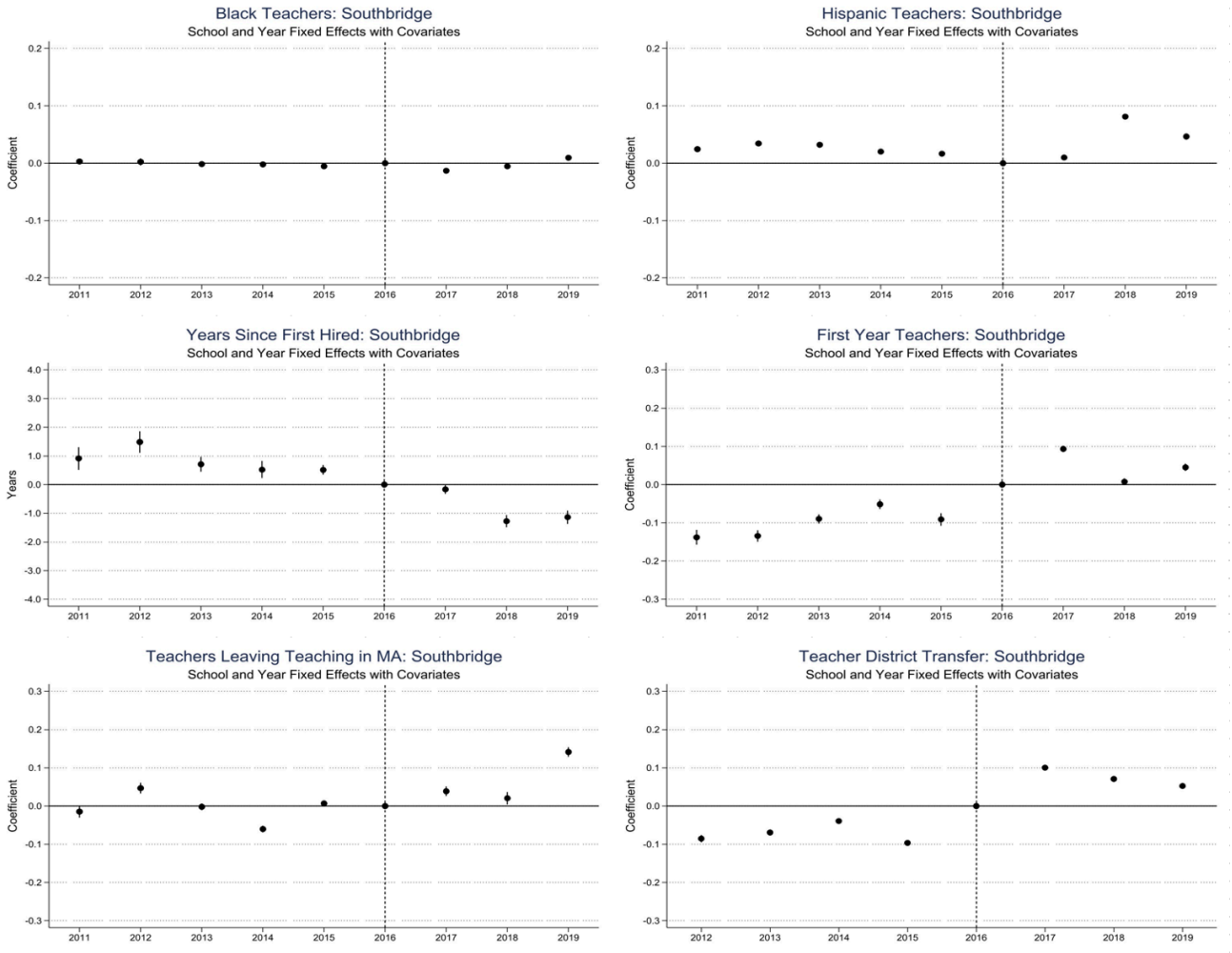

Figure 16. Southbridge reform impacts on teacher characteristics, event study estimates.
Note: The turnover variables are backward-looking (e.g., turnover rates for 2013 reflect whether teachers who were present in 2011-12 were also present in 2012-13).

Table 1. Baseline Characteristics of the Student Sample

	Lawrence	Holyoke	SEZP Cohort 1	SEZP Cohort 2	Southbridge	Rest of MA	Majority Low-Income Districts
Female	0.47	0.48	0.48	0.49	0.48	0.49	0.48
Low Income	0.92	0.84	0.91	0.90	0.73	0.33	0.75
White	0.07	0.18	0.14	0.09	0.48	0.70	0.32
Black	0.02	0.03	0.19	0.25	0.02	0.08	0.23
Hispanic	0.88	0.77	0.60	0.62	0.47	0.12	0.32
Asian	0.02	0.01	0.02	0.01	0.01	0.05	0.07
Other	0.01	0.01	0.05	0.03	0.02	0.04	0.06
First Language Not English	0.82	0.59	0.33	0.39	0.30	0.17	0.41
Limited English Proficient	0.40	0.37	0.27	0.31	0.19	0.10	0.26
SPED	0.28	0.38	0.31	0.30	0.28	0.26	0.27
Math Score	-0.72	-0.82	-0.90	-1.20	-0.60	0.03	-0.42
ELA Score	-0.73	-0.95	-0.92	-1.12	-0.67	0.03	-0.48
N of Districts	1	1	1	1	1	395	54
N of Students	12,278	5,194	4,352	1,371	2,059	866,575	214,232

Note: For each of the turnaround contexts, we average across all pre-turnaround years. For example, for Lawrence, we average across 2007-08 through 2011-12. For the comparison groups in the last two columns, we average across the years that represent pre-turnaround years for all districts (2007-08 through 2011-12). The N of Students represents the number of students in a single pre-turnaround year.

Table 2. Baseline Characteristics of the Teacher Sample

	Lawrence	Holyoke	SEZP Cohort 1	SEZP Cohort 2	Southbridge	Rest of MA	Majority Low-Income Districts
Female	77.2	70.8	71.3	60.6	71.3	74.8	73.8
White	87.8	88.0	77.9	67.7	93.3	93.5	82.9
Black	1.2	1.5	10.6	14.8	1.4	2.6	8.3
Hispanic	9.6	9.6	8.1	13.5	4.0	2.0	5.7
Asian	0.8	0.5	1.7	3.3	0.7	1.2	2.3
Other	0.6	0.4	1.8	0.6	0.7	0.6	0.9
First Year MA Teacher	14.1	10.4	17.5	18.4	12.8	8.8	11.5
Years Since First Hired in MA	9.2	12.6	9.3	9.5	11.2	11.2	11.4
Departure From Teaching or MA	10.3	11.1	13.1	13.8	12.0	8.4	9.9
Transfer to Other MA District	3.2	5.5	17.0	16.7	7.4	3.1	3.6
N of Districts	1	1	1	1	1	403	60
N of Unique Teachers	1,276	882	871	338	381	86,372	27,343

Note: For each of the turnaround contexts, we average across all pre-turnaround years. For example, for Lawrence, we average across 2007-08 through 2011-12. For the comparison groups in the last two columns, we average across the years that represent pre-turnaround years for all districts (2007-08 through 2011-12). The N of teachers represents the total number of in-person teachers and co-teachers in the given pre-years. The Years Since First Hired in MA variable is based on the first time an employee was hired by the MA public school system, regardless of whether they were hired as a teacher or another position and regardless of whether the person took breaks in employment. The First Year Observed MA Teacher is based on whether it is the first year a person was identified in the data as an in-person teacher or co-teacher in MA (all 2008 observations are missing as the data are censored prior to that year).

Table 3. The Effect of Lawrence Reforms, Event Study Estimates

	Math	ELA	Science	Days Attended	In-School Suspensions	Out-of-School Suspensions	Grade Progression
Lawrence 2007	-0.04*	-0.05*	-0.10***	-3.43***	0.01**	-0.03*	-0.02***
	(0.02)	(0.02)	(0.02)	(0.74)	(0.00)	(0.02)	(0.00)
Lawrence 2008	-0.03	-0.00	-0.01	-2.54	0.01	-0.03**	-0.05***
	(0.02)	(0.02)	(0.02)	(1.37)	(0.01)	(0.01)	(0.00)
Lawrence 2009	-0.02	-0.01	-0.00	-1.73***	0.00	-0.01	-0.01***
	(0.01)	(0.02)	(0.01)	(0.37)	(0.00)	(0.01)	(0.00)
Lawrence 2010	0.08***	0.10***	0.01	-1.46	0.04***	0.02*	-0.01*
	(0.01)	(0.01)	(0.01)	(1.05)	(0.00)	(0.01)	(0.00)
Lawrence 2011	0.00	0.03***	-0.01	-4.88***	0.03***	0.02	-0.00
	(0.01)	(0.01)	(0.01)	(0.43)	(0.00)	(0.01)	(0.00)
Lawrence 2013	0.16***	-0.01	-0.01	1.55***	-0.07**	-0.10***	0.01***
	(0.01)	(0.01)	(0.01)	(0.43)	(0.02)	(0.01)	(0.00)
Lawrence 2014	0.25***	0.03***	0.04*	2.19***	-0.05**	-0.09***	0.01***
	(0.01)	(0.01)	(0.01)	(0.53)	(0.02)	(0.01)	(0.00)
Lawrence 2015	0.27***	0.04**	0.13***	0.20	-0.06***	-0.06**	0.00
	(0.03)	(0.01)	(0.02)	(0.66)	(0.01)	(0.02)	(0.00)
Lawrence 2016	0.20***	-0.01	0.12***	0.38	-0.05***	-0.01	0.01***
	(0.03)	(0.02)	(0.03)	(1.01)	(0.01)	(0.02)	(0.00)
Lawrence 2017	0.27***	0.08***	0.21***	-2.31**	-0.05***	-0.01	0.01**
	(0.04)	(0.02)	(0.03)	(0.76)	(0.01)	(0.02)	(0.00)
Lawrence 2018	0.19***	0.03	0.21***	-2.07**	-0.03**	-0.00	-0.01
	(0.04)	(0.02)	(0.03)	(0.65)	(0.01)	(0.01)	(0.00)
Lawrence 2019	0.12**	-0.04	0.10***	-3.48***	-0.03***	-0.02	-
	(0.04)	(0.03)	(0.03)	(0.56)	(0.01)	(0.02)	-
Observations	1,487,975	1,485,905	796,348	2,926,570	2,998,701	2,998,701	2,761,226
Joint F-Test	0.03	0.84	3.68	17.55	49.49	0.72	40.12
	[0.87]	[0.36]	[0.06]	[0.00]	[0.00]	[0.40]	[0.00]

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 4. Variation in the Effect of Lawrence Reforms by Student Demographic Characteristics

	Average Pooled Effect	Black	Hispanic	Low Income	FLNE	SPED	Immigrant
				Math			
Lawrence Post	0.21***	0.21***	0.09***	-0.10*	0.21***	0.22***	0.23***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Lawrence Post X Characteristic		0.08***	0.13***	0.33***	0.00	-0.05*	-0.11***
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	1,487,975	1,487,975	1,487,975	1,487,975	1,487,975	1,487,975	1,487,975
				ELA			
Lawrence Post	0.00	0.00	-0.11***	-0.34***	-0.02	0.00	0.03
	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)
Lawrence Post X Characteristic		0.06*	0.12***	0.37***	0.02	-0.00	-0.11***
		(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Observations	1,485,905	1,485,905	1,485,905	1,485,905	1,485,905	1,485,905	1,485,905
				Science			
Lawrence Post	0.12***	0.12***	0.03	-0.14**	0.16***	0.12***	0.15***
	(0.02)	(0.02)	(0.03)	(0.05)	(0.04)	(0.03)	(0.03)
Lawrence Post X Characteristic		0.02	0.10***	0.28***	-0.05*	0.01	-0.11***
		(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Observations	796,348	796,348	796,348	796,348	796,348	796,348	796,348
				Days of Attendance			
Lawrence Post	1.78***	1.83***	-1.59*	-4.34***	1.04	1.27*	1.71***
	(0.43)	(0.43)	(0.60)	(1.22)	(0.84)	(0.48)	(0.45)
Lawrence Post X Characteristic		-3.52***	3.83***	6.79***	0.97	1.99***	0.31
		(0.23)	(0.30)	(1.07)	(0.60)	(0.37)	(0.63)
Observations	2,926,570	2,926,570	2,926,570	2,926,570	2,926,570	2,926,570	2,926,570
				In-School Suspensions			
Lawrence Post	-0.06***	-0.06***	-0.06***	-0.04***	-0.06***	-0.06***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Lawrence Post X Characteristic		-0.01*	-0.01*	-0.02*	-0.00	-0.01	0.00
		(0.01)	(0.00)	(0.01)	(0.00)	(0.02)	(0.01)
Observations	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701
				Out-of-School Suspensions			
Lawrence Post	-0.03***	-0.03**	-0.03*	0.00	-0.05***	-0.02	-0.04***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Lawrence Post X Characteristic		-0.04***	-0.01	-0.04*	0.02***	-0.04	0.02
		(0.01)	(0.01)	(0.02)	(0.00)	(0.03)	(0.01)
Observations	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701
				Grade Progression			
Lawrence Post	0.02***	0.02***	-0.00	-0.03***	0.01	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Lawrence Post X Characteristic		-0.02***	0.03***	0.06***	0.02***	0.00*	0.02***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Observations	2,761,226	2,761,226	2,761,226	2,761,226	2,761,226	2,761,226	2,761,226

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 5. Variation in the Effect of Reforms by Student Grade Level

	Math	ELA	Science	Attendance	In-School Suspensions	Out-of-School Suspensions	Grade Progression
				Lawrence			
Lawrence Post Elementary	0.07*	-0.08**	0.07	0.47	-0.01***	-0.01	0.00*
	(0.03)	(0.02)	(0.04)	(0.29)	(0.00)	(0.01)	(0.00)
Lawrence Post X Middle	0.25***	0.14***	0.11***	0.10	-0.04***	-0.07***	-0.02***
	(0.02)	(0.01)	(0.03)	(0.29)	(0.01)	(0.01)	(0.00)
Lawrence Post X High	0.15***	0.13***	0.05	4.27***	-0.14***	-0.03	0.08***
	(0.03)	(0.02)	(0.03)	(0.78)	(0.03)	(0.02)	(0.01)
Observations	1,487,975	1,485,905	796,348	2,926,570	2,998,701	2,998,701	2,761,226
				Holyoke			
Holyoke Post Elementary	-0.22***	-0.03	-0.12***	0.42	-0.01*	0.06***	0.01***
	(0.02)	(0.02)	(0.02)	(0.44)	(0.00)	(0.01)	(0.00)
Holyoke Post X Middle	0.00	-0.03	0.08**	0.51*	-0.07***	-0.03*	-0.02***
	(0.02)	(0.02)	(0.02)	(0.25)	(0.00)	(0.01)	(0.00)
Holyoke Post X High	-0.04	-0.11***	0.11***	-0.49	0.03	-0.37***	0.02***
	(0.02)	(0.02)	(0.02)	(0.77)	(0.02)	(0.01)	(0.00)
Observations	1,154,571	1,152,471	620,630	2,190,931	2,243,038	2,243,038	2,013,908
				Southbridge			
Southbridge Elementary	-0.32***	-0.22***	-0.21***	-2.01**	-0.00*	0.03***	0.01**
	(0.02)	(0.02)	(0.01)	(0.58)	(0.00)	(0.01)	(0.00)
Southbridge X Middle	0.17***	0.13***	0.01	0.70	0.11***	0.11***	-0.01*
	(0.02)	(0.01)	(0.02)	(0.36)	(0.01)	(0.01)	(0.00)
Southbridge X High	0.21***	-0.00	-0.22***	1.79	0.02	-0.07***	-0.01*
	(0.02)	(0.02)	(0.02)	(0.91)	(0.02)	(0.01)	(0.00)
Observations	1,031,198	1,028,823	553,029	1,951,250	1,996,627	1,996,627	1,770,369

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 6. The Effect of Holyoke Reforms, Event Study Estimates

	Math	ELA	Science	Days Attended	In-School Suspensions	Out-of-School Suspensions	Grade Progression
Holyoke 2010	0.05	0.06**	-0.02	-4.20***	0.05***	0.25***	-0.01*
	(0.03)	(0.02)	(0.02)	(0.84)	(0.01)	(0.02)	(0.00)
Holyoke 2011	0.09**	0.07***	-0.05*	-2.77***	0.05***	0.21***	0.01**
	(0.03)	(0.01)	(0.02)	(0.73)	(0.01)	(0.02)	(0.00)
Holyoke 2012	0.04	0.06***	-0.05*	-2.99***	0.04***	0.16***	0.01***
	(0.03)	(0.01)	(0.02)	(0.84)	(0.01)	(0.02)	(0.00)
Holyoke 2013	0.09***	0.02	-0.06**	-3.68***	-0.01	0.54***	0.01**
	(0.02)	(0.01)	(0.02)	(0.91)	(0.03)	(0.02)	(0.00)
Holyoke 2014	0.04*	-0.07***	-0.09***	-1.54*	-0.02	0.42***	0.01***
	(0.02)	(0.01)	(0.01)	(0.71)	(0.01)	(0.02)	(0.00)
Holyoke 2016	-0.05*	-0.02	-0.10***	-1.28*	-0.00	0.15***	0.03***
	(0.02)	(0.02)	(0.01)	(0.54)	(0.01)	(0.01)	(0.00)
Holyoke 2017	-0.20***	-0.05***	-0.14***	-1.09	0.00	0.16***	0.01***
	(0.02)	(0.01)	(0.02)	(0.60)	(0.01)	(0.01)	(0.00)
Holyoke 2018	-0.19***	-0.02	-0.08***	-2.46**	-0.01	0.24***	0.04***
	(0.02)	(0.02)	(0.02)	(0.73)	(0.01)	(0.01)	(0.00)
Holyoke 2019	-0.26***	-0.09**	-0.01	-3.81***	0.05***	0.15***	-
	(0.03)	(0.03)	(0.02)	(0.74)	(0.01)	(0.01)	-
Observations	1,154,571	1,152,471	620,630	2,190,931	2,243,038	2,243,038	2,013,908
Joint F-Test	6.40	4.81	9.65	22.74	6.57	336.92	10.26
	[0.01]	[0.03]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 7. Variation in the Effect of Holyoke Reforms by Student Demographic Characteristics

	Average Pooled Effect	Black	Hispanic	Low Income	FLNE	SPED	Immigrant
				Math			
Holyoke Post	-0.22***	-0.22***	-0.14***	-0.18***	-0.17***	-0.29***	-0.22***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Holyoke Post X Characteristic		0.05*	-0.11***	-0.04*	-0.10***	0.21***	0.24***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Observations	1,154,571	1,154,571	1,154,571	1,154,571	1,154,571	1,154,571	1,154,571
				ELA			
Holyoke Post	-0.06***	-0.07***	0.07**	-0.01	-0.00	-0.13***	-0.07***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Holyoke Post X Characteristic		0.06*	-0.16***	-0.06**	-0.12***	0.20***	0.16***
		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Observations	1,152,471	1,152,471	1,152,471	1,152,471	1,152,471	1,152,471	1,152,471
				Science			
Holyoke Post	-0.04*	-0.05*	0.11**	0.13*	0.00	-0.10**	-0.05*
	(0.02)	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)	(0.02)
Holyoke Post X Characteristic		0.05*	-0.19***	-0.20***	-0.08***	0.14***	0.42***
		(0.02)	(0.03)	(0.04)	(0.02)	(0.04)	(0.04)
Observations	620,630	620,630	620,630	620,630	620,630	620,630	620,630
				Days of Attendance			
Holyoke Post	0.36	0.56	1.47*	4.39***	2.79***	0.40	0.43
	(0.53)	(0.53)	(0.62)	(0.90)	(0.61)	(0.45)	(0.52)
Holyoke Post X Characteristic		-7.21***	-1.38*	-4.58***	-4.68***	-0.12	-17.50***
		(0.22)	(0.66)	(1.15)	(0.77)	(0.43)	(0.99)
Observations	2,190,931	2,190,931	2,190,931	2,190,931	2,190,931	2,190,931	2,190,931
				In-School Suspensions			
Holyoke Post	-0.01	-0.01	0.00	0.01	-0.00	-0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Holyoke Post X Characteristic		-0.01	-0.02**	-0.02*	-0.02***	-0.03*	-0.03**
		(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Observations	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038
				Out-of-School Suspensions			
Holyoke Post	-0.08***	-0.08***	-0.11***	-0.15***	-0.07***	-0.08***	-0.08***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Holyoke Post X Characteristic		0.06***	0.04***	0.08***	-0.02**	0.01	-0.02
		(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Observations	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038
				Grade Progression			
Holyoke Post	0.02***	0.02***	0.02***	0.04***	0.03***	0.01***	0.02***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Holyoke Post X Characteristic		-0.04***	-0.01	-0.03**	-0.02***	0.02***	-0.10***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
Observations	2,013,908	2,013,908	2,013,908	2,013,908	2,013,908	2,013,908	2,013,908

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 8. The Effect of SEZP Cohort 1 Reforms, Event Study Estimates

	Math	ELA	Science	Days Attended	In-School Suspensions	Out-of-School Suspensions	Grade Progression
SEZP 1 2010	-0.04	0.15***	0.19***	-3.26***	0.02	-0.24***	0.02***
	(0.04)	(0.02)	(0.03)	(0.80)	(0.01)	(0.02)	(0.00)
SEZP 1 2011	-0.06	0.08***	0.11***	-4.35***	0.00	-0.23***	0.01**
	(0.03)	(0.01)	(0.02)	(0.70)	(0.01)	(0.02)	(0.00)
SEZP 1 2012	-0.08*	0.05**	0.09**	-3.48***	-0.00	-0.13***	0.02***
	(0.03)	(0.02)	(0.03)	(0.74)	(0.01)	(0.03)	(0.00)
SEZP 1 2013	-0.10**	0.03	0.09***	-1.03	0.08***	-0.01	0.01***
	(0.04)	(0.01)	(0.02)	(0.87)	(0.01)	(0.02)	(0.00)
SEZP 1 2014	-0.02	0.03*	-0.05**	0.19	0.09***	-0.15***	0.01**
	(0.02)	(0.01)	(0.02)	(0.40)	(0.01)	(0.02)	(0.00)
SEZP 1 2016	-0.00	0.06***	0.00	-1.15*	0.02	0.00	0.02***
	(0.02)	(0.02)	(0.02)	(0.55)	(0.01)	(0.01)	(0.00)
SEZP 1 2017	0.14***	0.24***	0.06**	-1.75*	0.01	-0.09***	0.03***
	(0.03)	(0.03)	(0.02)	(0.67)	(0.01)	(0.02)	(0.00)
SEZP 1 2018	0.13***	0.24***	0.12***	-0.90	-0.05***	-0.06**	0.01*
	(0.04)	(0.04)	(0.02)	(0.70)	(0.01)	(0.02)	(0.00)
SEZP 1 2019	0.18***	0.27***	0.18***	-2.00**	-0.05***	-0.13***	-
	(0.05)	(0.05)	(0.03)	(0.58)	(0.01)	(0.01)	-
Observations	1,158,630	1,156,626	600,143	2,136,791	2,185,566	2,185,566	1,961,175
Joint F-Test	3.59	26.2	18.85	15.4	15.89	52.72	27.5
	[0.06]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 9. The Effect of SEZP Cohort 2 Reforms, Event Study Estimates

	Math	ELA	Science	Days Attended	In-School Suspensions	Out-of-School Suspensions	Grade Progression
SEZP 2 2012	-0.30***	0.12***	-0.18***	-18.15***	0.19***	0.04	-0.10***
	(0.03)	(0.02)	(0.04)	(1.16)	(0.02)	(0.03)	(0.01)
SEZP 2 2013	-0.23***	0.15***	-0.23***	-18.82***	0.38***	0.01	-0.14***
	(0.03)	(0.02)	(0.03)	(0.92)	(0.07)	(0.05)	(0.01)
SEZP 2 2014	0.13***	0.29***	-0.04	-8.73***	0.17***	-0.10*	-0.01
	(0.02)	(0.02)	(0.02)	(0.90)	(0.04)	(0.05)	(0.01)
SEZP 2 2015	0.07**	0.16***	-0.05*	-3.81***	-0.06**	0.27***	-0.01**
	(0.02)	(0.02)	(0.02)	(0.55)	(0.02)	(0.01)	(0.00)
SEZP 2 2016	0.10***	0.09***	-0.15***	-0.00	-0.03***	-0.06***	-0.03***
	(0.02)	(0.01)	(0.01)	(0.64)	(0.01)	(0.01)	(0.01)
SEZP 2 2018	0.08***	0.03	0.15***	-0.91*	0.02	-0.02	-0.02***
	(0.02)	(0.02)	(0.02)	(0.41)	(0.03)	(0.02)	(0.00)
SEZP 2 2019	0.44***	0.59***	0.41***	1.78**	-0.01	-0.01	-
	(0.06)	(0.06)	(0.03)	(0.66)	(0.03)	(0.02)	-
Observations	913,893	911,451	492,574	1,734,416	1,775,888	1,775,888	1,550,392
Joint F-Test	4.36	120.76	32.71	246.06	51.56	1.65	194.95
	[0.04]	[0.00]	[0.00]	[0.00]	[0.00]	[0.20]	[0.00]

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 10. Variation in the Effect of SEZP Cohort 1 Reforms by Student Demographic Characteristics

	Average Pooled Effect	Black	Hispanic	Low Income	FLNE	SPED	Immigrant
				Math			
SEZP Coh. 1 Post	0.16**	0.15**	0.22***	0.22***	0.20***	0.16**	0.17**
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
SEZP Coh. 1 Post X Characteristic		0.07**	-0.09***	-0.07**	-0.11***	0.02	-0.06
		(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)
Observations	1,158,630	1,158,630	1,158,630	1,158,630	1,158,630	1,158,630	1,158,630
				ELA			
SEZP Coh. 1 Post	0.14***	0.14***	0.19***	0.20***	0.18***	0.15***	0.15***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
SEZP Coh. 1 Post X Characteristic		0.05	-0.07***	-0.06*	-0.11***	-0.01	-0.07
		(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
Observations	1,156,626	1,156,626	1,156,626	1,156,626	1,156,626	1,156,626	1,156,626
				Science			
SEZP Coh. 1 Post	0.02	0.02	0.06*	0.03	0.05	-0.01	0.03
	(0.03)	(0.03)	(0.03)	(0.05)	(0.03)	(0.03)	(0.03)
SEZP Coh. 1 Post X Characteristic		0.02	-0.06***	-0.01	-0.08***	0.12***	-0.09**
		(0.02)	(0.02)	(0.04)	(0.02)	(0.03)	(0.03)
Observations	600,143	600,143	600,143	600,143	600,143	600,143	600,143
				Days of Attendance			
SEZP Coh. 1 Post	0.58	0.69	0.11	2.55***	1.32**	1.23**	0.53
	(0.47)	(0.48)	(0.41)	(0.63)	(0.40)	(0.42)	(0.46)
SEZP Coh. 1 Post X Characteristic		-0.64**	0.69**	-2.12*	-2.10**	-2.18***	0.93
		(0.20)	(0.25)	(0.95)	(0.62)	(0.37)	(0.76)
Observations	2,136,791	2,136,791	2,136,791	2,136,791	2,136,791	2,136,791	2,136,791
				In-School Suspensions			
SEZP Coh. 1 Post	-0.05***	-0.05***	-0.05***	-0.05***	-0.05***	-0.04***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SEZP Coh. 1 Post X Characteristic		0.01	-0.01	0.00	-0.01***	-0.02	-0.01
		(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Observations	2,185,566	2,185,566	2,185,566	2,185,566	2,185,566	2,185,566	2,185,566
				Out-of-School Suspensions			
SEZP Coh. 1 Post	0.06**	0.06**	0.05*	-0.09**	0.10***	0.02	0.07***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
SEZP Coh. 1 Post X Characteristic		0.03***	0.02*	0.17***	-0.12***	0.13***	-0.22***
		(0.01)	(0.01)	(0.01)	(0.00)	(0.03)	(0.01)
Observations	2,185,566	2,185,566	2,185,566	2,185,566	2,185,566	2,185,566	2,185,566
				Grade Progression			
SEZP Coh. 1 Post	0.01**	0.01***	-0.00	-0.01	0.01**	0.00	0.01**
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
SEZP Coh. 1 Post X Characteristic		-0.02***	0.02***	0.02**	-0.01	0.02***	-0.02**
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Observations	1,961,175	1,961,175	1,961,175	1,961,175	1,961,175	1,961,175	1,961,175

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 11. Variation in the Effect of SEZP Cohort 2 Reforms by Student Demographic Characteristics

	Average Pooled Effect	Black	Hispanic	Low Income	FLNE	SPED	Immigrant
				Math			
SEZP Coh. 2 Post	0.28***	0.26***	0.44***	0.48***	0.41***	0.26***	0.36***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
SEZP Coh. 2 Post X Characteristic		0.13***	-0.24***	-0.21***	-0.30***	0.10***	-0.51***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Observations	913,893	913,893	913,893	913,893	913,893	913,893	913,893
				ELA			
SEZP Coh. 2 Post	0.19***	0.14***	0.37***	0.35***	0.45***	0.12***	0.35***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
SEZP Coh. 2 Post X Characteristic		0.26***	-0.26***	-0.17***	-0.61***	0.26***	-1.04***
		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Observations	911,451	911,451	911,451	911,451	911,451	911,451	911,451
				Science			
SEZP Coh. 2 Post	0.39***	0.37***	0.53***	0.48***	0.47***	0.41***	0.44***
	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)
SEZP Coh. 2 Post X Characteristic		0.07**	-0.20***	-0.11*	-0.20***	-0.10*	-0.49***
		(0.03)	(0.02)	(0.04)	(0.03)	(0.04)	(0.04)
Observations	492,574	492,574	492,574	492,574	492,574	492,574	492,574
				Days of Attendance			
SEZP Coh. 2 Post	8.16***	7.35***	9.05***	9.81***	10.05***	9.27***	7.75***
	(0.61)	(0.62)	(0.70)	(0.98)	(0.76)	(0.60)	(0.59)
SEZP Coh. 2 Post X Characteristic		4.25***	-1.30***	-1.76	-4.16***	-3.57***	3.99***
		(0.27)	(0.34)	(0.89)	(0.66)	(0.34)	(0.69)
Observations	1,734,416	1,734,416	1,734,416	1,734,416	1,734,416	1,734,416	1,734,416
				In-School Suspensions			
SEZP Coh. 2 Post	-0.09**	-0.09**	-0.08*	-0.06	-0.09**	-0.08*	-0.09**
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
SEZP Coh. 2 Post X Characteristic		0.00	-0.02***	-0.03**	-0.00	-0.04***	0.01
		(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Observations	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888
				Out-of-School Suspensions			
SEZP Coh. 2 Post	-0.05**	-0.04*	-0.02	-0.14***	-0.02	-0.06*	-0.03*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)
SEZP Coh. 2 Post X Characteristic		-0.03***	-0.04***	0.10***	-0.05***	0.03	-0.11***
		(0.01)	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)
Observations	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888
				Grade Progression			
SEZP Coh. 2 Post	0.02***	0.01**	0.08***	-0.08***	0.04***	0.02***	0.01***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
SEZP Coh. 2 Post X Characteristic		0.05***	-0.09***	0.11***	-0.04***	-0.01***	0.05***
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Observations	1,550,392	1,550,392	1,550,392	1,550,392	1,550,392	1,550,392	1,550,392

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 12. The Effect of Southbridge Reforms, Event Study Estimates

	Math	ELA	Science	Days Attended	In-School Suspensions	Out-of-School Suspensions	Grade Progression
Southbridge 2011	0.25***	0.27***	0.38***	3.65**	0.04**	-0.13***	-0.01
	(0.03)	(0.02)	(0.03)	(1.34)	(0.01)	(0.01)	(0.00)
Southbridge 2012	0.17***	0.20***	0.34***	3.94*	-0.00	-0.12***	0.00
	(0.03)	(0.02)	(0.03)	(1.56)	(0.01)	(0.02)	(0.00)
Southbridge 2013	0.14***	0.08***	0.11***	3.56*	0.01	0.01	-0.01
	(0.03)	(0.02)	(0.02)	(1.51)	(0.02)	(0.02)	(0.00)
Southbridge 2014	0.18***	0.13***	0.10***	2.96*	0.07***	-0.02	0.01***
	(0.02)	(0.01)	(0.02)	(1.36)	(0.01)	(0.02)	(0.00)
Southbridge 2015	0.14***	0.10***	0.06***	4.57***	0.04***	0.00	-0.05***
	(0.02)	(0.02)	(0.01)	(1.21)	(0.01)	(0.01)	(0.00)
Southbridge 2017	0.05**	0.08***	-0.12***	2.62**	0.00	0.02***	-0.02***
	(0.02)	(0.01)	(0.01)	(0.93)	(0.00)	(0.00)	(0.00)
Southbridge 2018	-0.06***	-0.07***	-0.13***	2.61*	0.02	-0.04***	0.01***
	(0.02)	(0.02)	(0.02)	(1.06)	(0.01)	(0.01)	(0.00)
Southbridge 2019	-0.23***	-0.14***	-0.18***	-0.02	0.15***	0.03***	
	(0.02)	(0.02)	(0.02)	(1.11)	(0.01)	(0.01)	
Observations	1,031,198	1,028,823	553,029	1,951,250	1,996,627	1,996,627	1,770,369
Joint F-Test	57.06	93.88	91.15	7.47	41.17	13.61	16.21
	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

Table 13. Variation in the Effect of Southbridge Reforms by Student Demographic Characteristics

	Average Pooled Effect	Black	Hispanic	Low Income	FLNE	SPED	Immigrant
				Math			
Southbridge Post	-0.22***	-0.23***	-0.17***	-0.22***	-0.13***	-0.26***	-0.22***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Southbridge Post X Characteristic		0.39***	-0.08***	0.00	-0.28***	0.15***	-0.38***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Observations	1,031,198	1,031,198	1,031,198	1,031,198	1,031,198	1,031,198	1,031,198
				ELA			
Southbridge Post	-0.16***	-0.17***	-0.12***	-0.10***	-0.07***	-0.17***	-0.16***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Southbridge Post X Characteristic		0.29***	-0.09***	-0.08***	-0.29***	0.02	0.06*
		(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Observations	1,028,823	1,028,823	1,028,823	1,028,823	1,028,823	1,028,823	1,028,823
				Science			
Southbridge Post	-0.29***	-0.29***	-0.25***	-0.17***	-0.22***	-0.28***	-0.29***
	(0.01)	(0.01)	(0.02)	(0.04)	(0.02)	(0.02)	(0.01)
Southbridge Post X Characteristic		0.32***	-0.07**	-0.14***	-0.19***	-0.01	0.12***
		(0.02)	(0.02)	(0.04)	(0.02)	(0.04)	(0.03)
Observations	553,029	553,029	553,029	553,029	553,029	553,029	553,029
				Days of Attendance			
Southbridge Post	-1.31**	-1.10*	-0.70	3.06***	0.07	-1.00**	-1.34**
	(0.43)	(0.43)	(0.41)	(0.61)	(0.43)	(0.36)	(0.42)
Southbridge Post X Characteristic		-10.02***	-1.13*	-5.34***	-4.02***	-1.01*	7.67***
		(0.18)	(0.52)	(1.00)	(0.78)	(0.41)	(0.68)
Observations	1,951,250	1,951,250	1,951,250	1,951,250	1,951,250	1,951,250	1,951,250
				In-School Suspensions			
Southbridge Post	0.03***	0.03***	0.03**	0.00	0.03***	0.04***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Southbridge Post X Characteristic		-0.01*	0.01*	0.03**	0.01**	-0.01	0.14***
		(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Observations	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627
				Out-of-School Suspensions			
Southbridge Post	0.04***	0.05***	0.01	-0.00	0.03*	0.04*	0.04***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Southbridge Post X Characteristic		-0.18***	0.06***	0.05***	0.05***	0.00	-0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Observations	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627
				Grade Progression			
Southbridge Post	0.00	0.00	0.01**	0.01*	0.01***	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Southbridge Post X Characteristic		-0.03***	-0.01**	-0.01	-0.04***	0.01***	-0.16***
		(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
Observations	1,770,369	1,770,369	1,770,369	1,770,369	1,770,369	1,770,369	1,770,369

Note: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates and school-grade and grade-year fixed effects.

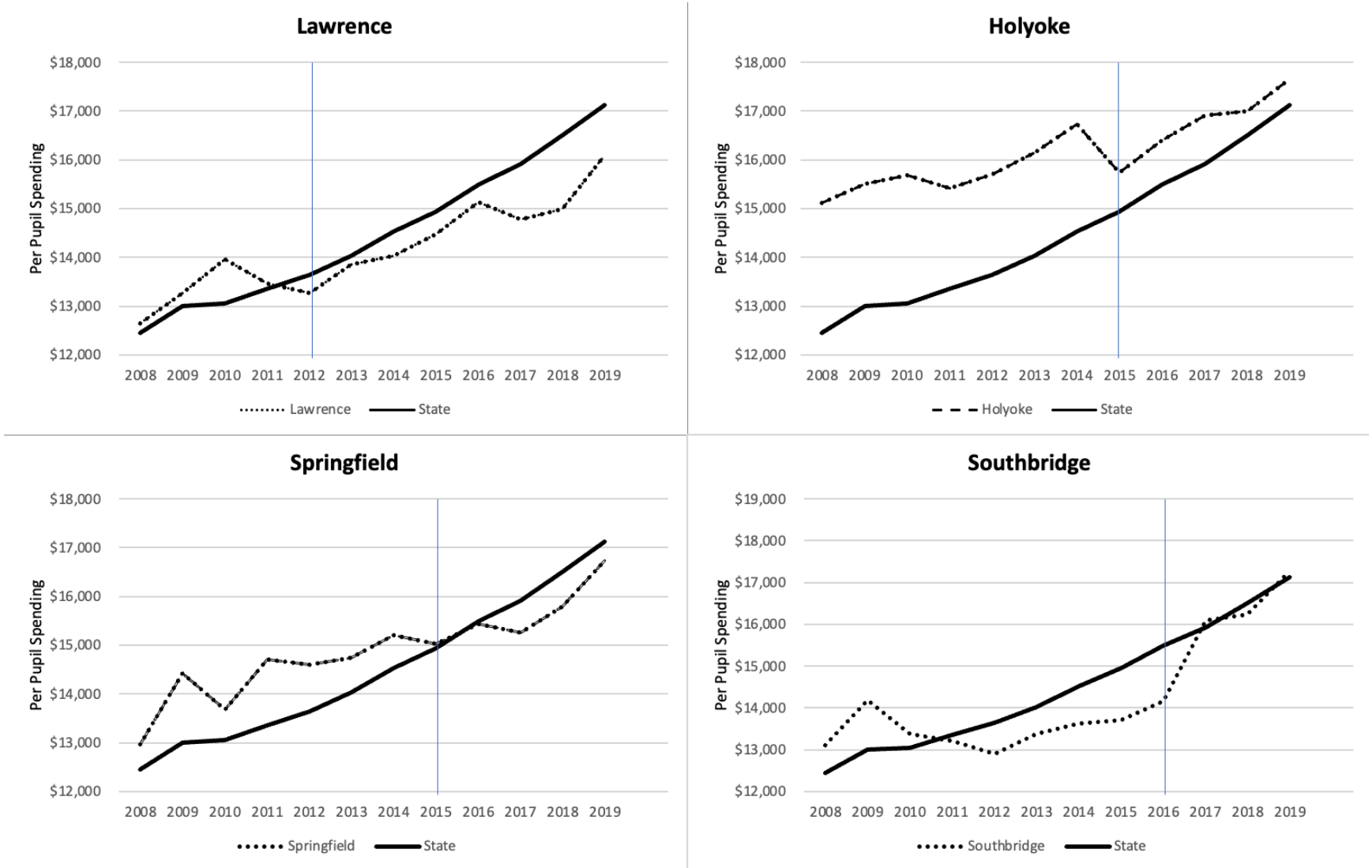

Figure A1. Average annual per pupil spending in turnaround districts versus Massachusetts as a whole (not adjusted for inflation). Note: Data publicly available from the Massachusetts Department of Elementary and Secondary Education. The Springfield figure represents all Springfield Public Schools (equivalent data not available for SEZP).

Lawrence							
	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
Autonomy & Accountability	Increased school autonomy at differentiated levels based on performance	Expanded principal autonomy over calendars, interim assessments, staffing; teacher leader teams influence school plans					Developed more explicit performance management framework for school autonomy and funding
		Central office budget reduced by 25%					Created structures for principal collaboration and capacity building
School Redesign	Gave management of one grade level at three schools to independent operators	Independent operators expanded to serve additional grades and schools					
	Independent operator opens new high school for students at risk of dropout	Union takes over management of one elementary school		Combined six high schools into one with pathways starting in grade 10			New principal at high school worked on bringing coherence across programs and improving climate
Learning Time	Ran Vacation Academies for 1800 students	Doubled Acceleration Academy participation and continued offering in all future years					
	MATCH provided high-dosage tutoring to 550 9-10th graders	School year expanded at least 200 hours for grades 1-8		Extended learning time provided to ninth graders through high school redesign	Some tutoring available across grade levels		Standardized the calendar districtwide
Human Capital	Replaced 36% of principals, 20% of APs, 10% of teachers	Replaced another 20% of principals					
		New teacher compensation system with career ladder, performance pay, stipends for ELT and teacher leadership; raised pay overall				Signed new contract, codifying previous reforms, increased transparency of career ladder selection process; raised pay overall	
Data Use	ANet provided training on data use for instructional improvement at nine	Anet expanded to work with 85% of K-8 schools					
Curriculum	Built out extracurricular offerings with community partners	Continued to build out extracurricular offerings further		Built out early college programs for high schoolers	Began shifting to vetted, standards-aligned curriculum districtwide, especially in science and ELA		
Family Engagement				Community partners connect families with 20 local employers	Established new family resource center		
School Climate							Required Superintendent approval for suspensions
							Began trying to earn buy-in for restorative justice approach to discipline

Figure A2. Components of the Lawrence turnaround reforms by year.
Note: changes listed in an earlier year continued through the later years unless otherwise noted.

Holyoke				
	2015-16	**2016-17**	**2017-18**	**2018-19**
Autonomy & Accountability	Increased school autonomy although some things remained standardized (e.g., start/end times)			
	New central office cabinet, built out central office school support team and principal capacity building			
School Redesign	Merged high schools onto one campus and created menu of pathways, enhanced career program and early college program	Increased availability of high school advanced coursework		
	Added pre-K program at one elementary school and full-day Kindergarten	Added pre-K program at two more elementary schools	Added pre-K program at a fourth school (>100 slots total added over 3 years)	
				Gave management of one middle school to charter operator and created STEM middle school
Learning Time	Offered Vacation Academies for 300 students grades 4-6, continued in different grades going forward	Extended school day by two hours for grades K-8		Expanded access to extracurriculars
Human Capital	Principal replacements; Replaced 16% of teachers		New teacher compensation system with career ladder, performance pay, stipends for ELT and teacher leadership; raised pay overall	
	Launched teacher recruitment initiative with TNTP, DESE, charter network and SEZP			
Data Use	Introduced districtwide formative assessments			
Curriculum	Expanded dual language program in two K classes	Expanded dual language by one grade per year going forward		
	Built enrichment offerings into the school day (e.g., arts, sports)		Instroduced technology-enhanced personalized learning in some middle school classes	Expanded middle school technology-enhanced personalized learning program
	Began introducing new curricular materials in math	Began introducing new curricular materials in ELA		
Family Engagement	Launched advisory groups for parents, students, teachers. New phone communication system and enhanced translation for families. Began administering climate surveys.		Expanded home-visiting program	
Facilities	Began addressing deferred maintenance and upgrading systems (e.g., internet)			

Figure A3. Components of the Holyoke turnaround reforms by year
Note: changes listed in an earlier year continued through the later years unless otherwise noted.

	SEZP			
	2015-16	**2016-17**	**2017-18**	**2018-19**
Autonomy & Accountability	Increased school autonomy at differentiated levels based on performance; principal and teacher leadership teams develop plans re: budget (80% decided by school), curriculum, staffing, schedule, and culture. District offers menu of services.			
	Hold school managers accountable for results. Use performance management framework to plan, tailor supports, and make leadership and reconfiguration decsisions.			
School Redesign		Reconfigured three middle schools, creating new schools, some managed by independent operators	High school added to SEZP and reconfigured into a new learning community	High school reconfigured again to add an additional high school model
Learning Time	Ran Vacation Academies for 7th graders	Expanded Vacation Academies to 500 students and continued offering throughout		
	Provided all students 90 minutes per week of small group tutoring	All schools extended learning time		
Human Capital	New teacher compensation system with career ladder, performance pay, stipends for ELT and teacher leadership; raised pay overall			
	Replaced 67% principals (over whole period) and ran principal training institute	Launched program to expand pipeline of school leaders		
	Launched teacher recruitment initiative with TNTP, DESE, charter network and Holyoke			
Data Use	Roadmap provides data for school- and zone-level planning			
	Many schools partner with Anet on data for instructional improvement			
Curriculum	Began shifting to new math and ELA curricula	Two new middle schools focused on dual language	New high school focused on early college	

Figure A4. Components of the SEZP turnaround reforms by year
Note: changes listed in an earlier year continued through the later years unless otherwise noted.

Southbridge			
	2016-17	**2017-18**	**2018-19**
Autonomy & Accountability	Focused on alignment across schools (with goal of increasing school autonomy in longer-run)		Created structures for principal collaboration and capacity building
School Redesign	New alternative middle / high school for students with behavioral issues; separated out middle and high school		Redesigned alternative school into therapeutic day school
	Separated out the middle and high school		
Learning Time	Extended learning time for elementary schools		
Human Capital	Principal and teacher review and replacements	New teacher compensation system with career ladder, performance pay, stipends for ELT; raised pay overall	Added one hour per week of teacher PD / common planning time
Data Use			Shifted from paper to digital record keeping systems for student information management, finance, facilities, operations, budget, food service, HR, etc.
			New data dashboards for principals
			Increased use of formative assessments and trained teachers
Curriculum	Increased number of instructional resource specialists	Began shifting to vetted, standards-aligned curriculum districtwide	New ELA curriculum in all grades; new math curriculum in grades 6-8
		Implemented dual language program in K-1st grade;	Expanded dual language program to grade 2
Family Engagement		Added family liaison at every school and new translation services for families	Hired director of communications
School Climate	Increased number of school psychologists and behavioral intervention specialists		Implemented PBIS at all schools; became national demonstration site for PBIS

Figure A5. Components of the Southbridge turnaround reforms by year.
Note: changes listed in an earlier year continued through the later years unless otherwise noted.

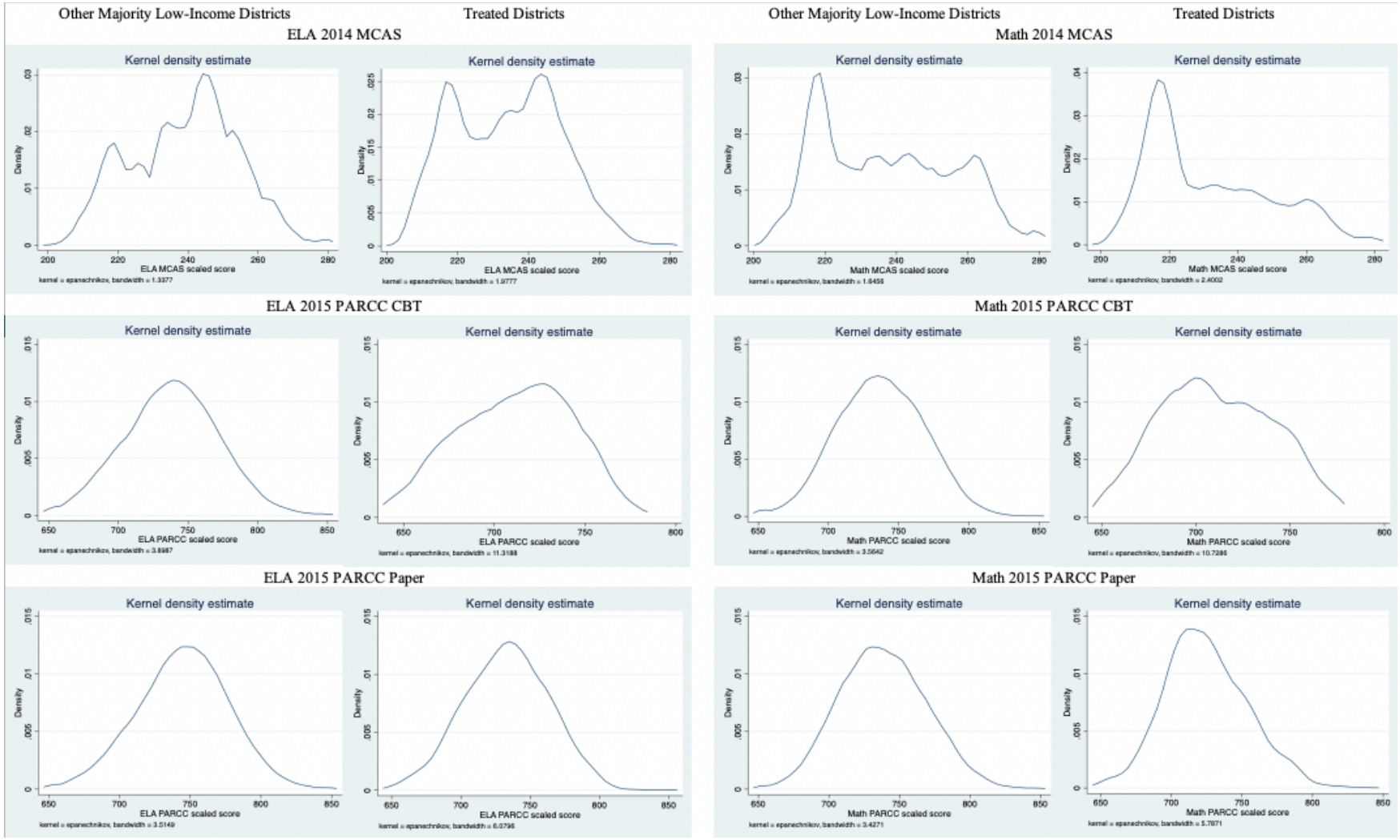

Figure A6. Kernel density plots of scaled scores (prior to standardization)

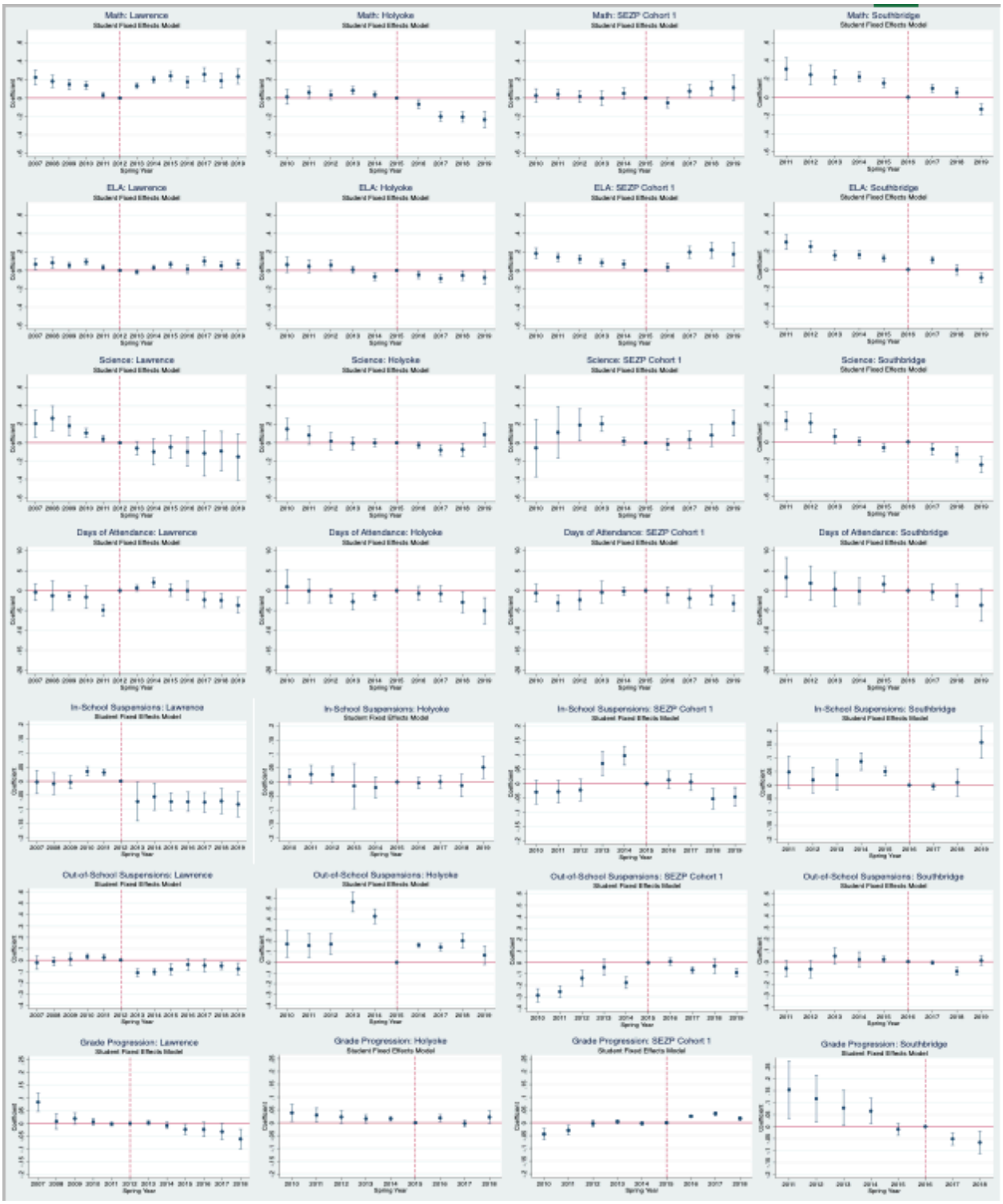

Figure A7. Student fixed effects estimates

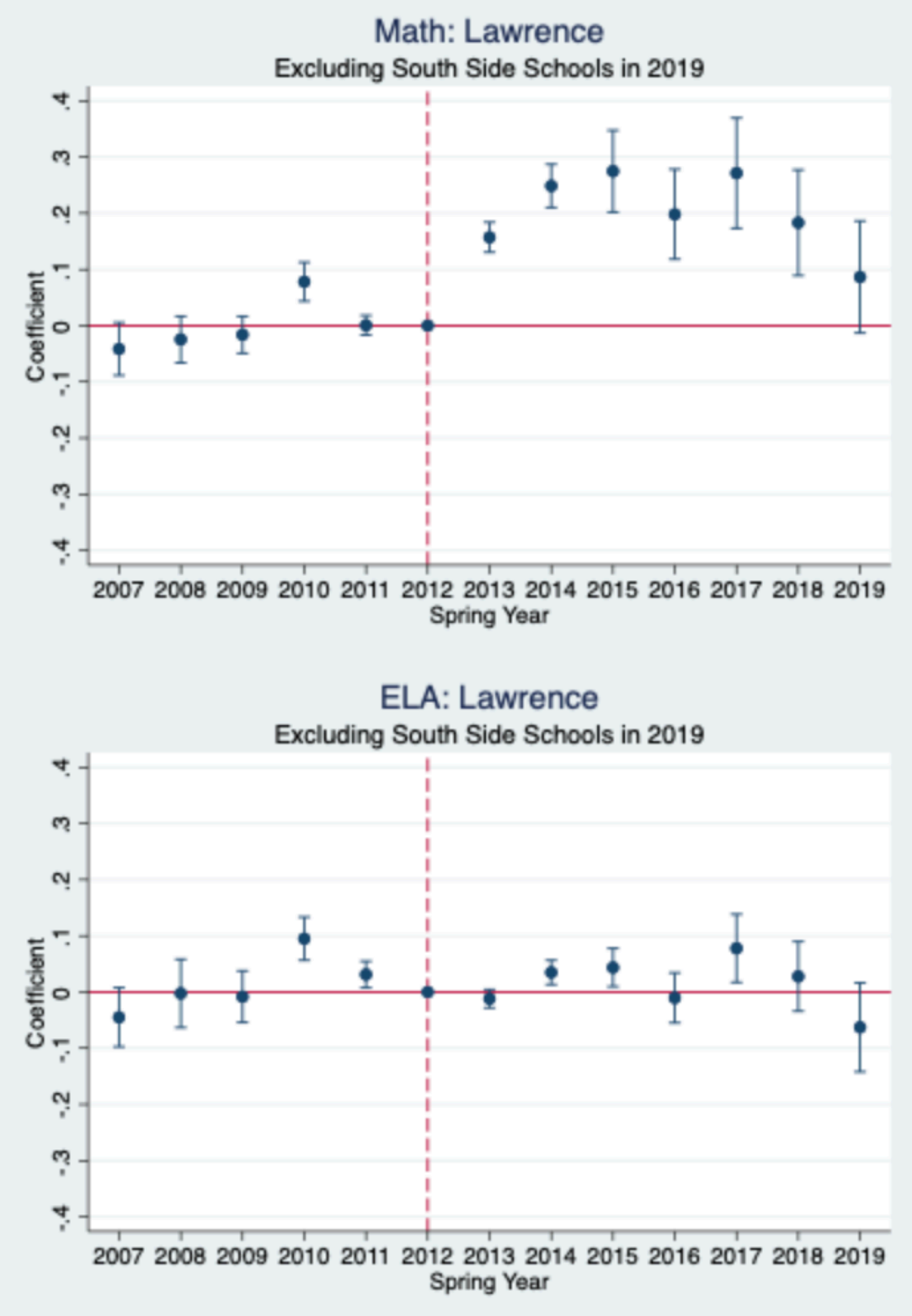

Figure A8. Effects of Lawrence reforms on test outcomes, excluding schools most impacted by gas explosions.

Appendix Table A1. The Effect of Reform on Demographic Composition of Treated Districts

	Asian	Black	Hispanic	Low Income	First Language Not English	Special Education	Immigrant
Lawrence Post Turnaround	-0.01**	-0.00	-0.05***	-0.07***	-0.10***	-0.01*	0.05***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701
Holyoke Post Turnaround	0.00	-0.00	-0.00	0.01	-0.08***	0.01	-0.02***
	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Observations	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038
SEZP Cohort 1 Post Turnaround	0.01	0.00	0.00	-0.02**	-0.02	0.01	0.02*
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Observations	2,235,768	2,235,768	2,235,768	2,235,768	2,235,768	2,235,768	2,235,768
SEZP Cohort 2 Post Turnaround	0.00	-0.02***	-0.01	-0.01*	-0.01	0.02***	0.02***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Observations	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888
Southbridge Post Turnaround	0.00*	0.01	0.01	0.05***	0.02	0.04**	-0.02**
	(0.00)	(0.00)	(0.02)	(0.01)	(0.05)	(0.01)	(0.01)
Observations	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627

Notes: *** p<0.001, ** p<0.01, * p<0.05. Each cell represents a separate regression.

Appendix Table A2. The Effect of Reform on Likelihood of Missing Outcomes

	Math	ELA	Science	Attendance	Grade Progression
Lawrence Post Turnaround	-1.65***	-1.22***	0.37	0.07	-0.00
	(0.31)	(0.25)	(0.52)	(0.19)	(0.00)
Observations	2,998,701	2,998,701	2,998,701	2,998,701	2,998,701
Holyoke Post Turnaround	-0.98	-0.91	1.97	-0.01	0.00
	(0.61)	(0.58)	(2.14)	(0.32)	(0.00)
Observations	2,243,038	2,243,038	2,243,038	2,243,038	2,243,038
SEZP Cohort 1 Post Turnaround	1.10	0.88	0.08	0.38	-0.00
	(1.25)	(1.24)	(0.30)	(0.32)	(0.00)
Observations	2,235,768	2,235,768	2,235,768	2,235,768	2,235,768
SEZP Cohort 2 Post Turnaround	0.26*	-0.06	-5.86***	5.41***	-0.00
	(0.11)	(0.12)	(1.14)	(0.16)	(0.00)
Observations	1,775,888	1,775,888	1,775,888	1,775,888	1,775,888
Southbridge Post Turnaround	0.28	-0.09	-2.35	-0.44	0.00
	(0.41)	(0.41)	(2.04)	(0.38)	(0.00)
Observations	1,996,627	1,996,627	1,996,627	1,996,627	1,996,627

Notes: *** p<0.001, ** p<0.01, * p<0.05. Each cell represents a separate regression. We do not calculate these estimates for the discipline outcomes as there is never any missingness on these outcomes.

Appendix Table A3. Pooled Effects of Reforms on Teacher Characteristics

	Black Teachers	Hispanic Teachers	Years Since First Hired	First Year Teachers	Depart State	Transfer District
Lawrence	0.00	0.01**	-0.40	0.04***	0.03***	0.06***
	(0.00)	(0.00)	(0.23)	(0.01)	(0.01)	(0.00)
Observations	182357	182357	183137	182357	182725	163211
Holyoke	0.02***	0.01**	-2.27***	0.09***	0.05***	0.11***
	(0.00)	(0.00)	(0.20)	(0.00)	(0.01)	(0.00)
Observations	160512	160398	160398	160398	160801	143437
SEZP Cohort 1	0.04***	-0.02***	-0.64**	0.00	0.01	0.06***
	(0.00)	(0.00)	(0.19)	(0.00)	(0.01)	(0.00)
Observations	159139	159139	159139	159139	159590	142386
SEZP Cohort 2	-0.02***	-0.00*	-1.13***	-0.03***	0.04***	-0.07***
	(0.00)	(0.00)	(0.18)	(0.00)	(0.01)	(0.00)
Observations	126160	126160	126160	126160	157813	140846
Southbridge	-0.00	0.02***	-1.84***	0.16***	0.07***	0.16***
	(0.00)	(0.00)	(0.20)	(0.01)	(0.01)	(0.00)
Observations	141843	141843	141843	141843	141576	126007

Notes: *** p<0.001, ** p<0.01, * p<0.05. All models include covariates.