ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-479

The Role of State Education Regulation: Evidence from the Texas Districts of Innovation Statute

Kylie L. Anglin
University of Connecticut

Traditional public schools in the United States must comply with a variety of regulations on educational inputs like teacher certification, maximum class sizes, and restrictions on staff contracts. Absent regulations, policymakers fear that troubled districts would make inappropriate decisions that would harm students. However, it is also possible that strict regulations hinder schools from optimizing student learning. This paper tests the salience of these two hypotheses within the context of a widespread deregulation effort in Texas which allows traditional public school districts to claim District of Innovation status and opt out of regulations not related to health, safety, and civil rights. Using a novel dataset of administration data merged with implementation information scraped from district websites, I estimate the impact of District of Innovation status with a difference-in-differences strategy where later implementers act as the comparison group for early implementers. I find that, despite the breadth of regulations exempted, regulatory autonomy does not significantly impact either math or reading achievement nor does it impact hiring or class sizes. Together, the results offer strong evidence against the hypothesis that regulation hinders school improvement and suggests that state input regulations play only a limited role in determining school decision-making or student achievement.

VERSION: October 2021

Suggested citation: Anglin, Kylie L. . (2021). The Role of State Education Regulation: Evidence from the Texas Districts of Innovation Statute. (EdWorkingPaper: 21-479). Retrieved from Annenberg Institute at Brown University: https://-doi.org/10.26300/pt72-sk75

The Role of State Education Regulation: Evidence from the Texas Districts of Innovation Statute

Kylie L. Anglin

University of Connecticut

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B140026 to the Rectors and Visitors of the University of Virginia as well as by the National Academy of Education/Spencer Dissertation Fellowship program. The opinions expressed are those of the authors and do not represent views of NAEd/Spencer, the Institute or the U.S. Department of Education. The author wishes to thank Vivian Wong, Beth Schueler, James Wyckoff, and Emily Wiseman for their feedback on earlier versions of this paper. All errors are those of the author.

Abstract

Traditional public schools in the United States must comply with a variety of regulations on educational inputs like teacher certification, maximum class sizes, and restrictions on staff contracts. Absent regulations, policymakers fear that troubled districts would make inappropriate decisions that would harm students. However, it is also possible that strict regulations hinder schools from optimizing student learning. This paper tests the salience of these two hypotheses within the context of a widespread deregulation effort in Texas which allows traditional public school districts to claim *District of Innovation* status and opt out of regulations not related to health, safety, and civil rights. Using a novel dataset of administration data merged with implementation information scraped from district websites, I estimate the impact of District of Innovation status with a difference-in-differences strategy where later implementers act as the comparison group for early implementers. I find that, despite the breadth of regulations exempted, regulatory autonomy does not significantly impact either math or reading achievement nor does it impact hiring or class sizes. Together, the results offer strong evidence against the hypothesis that regulation hinders school improvement and suggests that state input regulations play only a limited role in determining school decision-making or student achievement.

What is the role of state legislatures in improving public education? Is it to provide funding and standards for student outcomes and then step aside? Or do districts also need to be told *how* to educate students? In practice, states hold school districts accountable for academic achievement while also requiring them to prove compliance with a variety of regulations on educational inputs (Cohen et al., 2017). Commonly regulated inputs include minimum teacher qualifications, maximum class sizes, and requirements on instructional time (Education Commission of the States, 2005, 2017, 2019). Without these regulations, policymakers fear that troubled districts would make inappropriate decisions that would harm student learning (Fuhrman & Elmore, 1995). It is an open question whether this fear is well-grounded. If states do not regulate hiring and operations, it is possible that some schools may adopt low-quality inputs, thereby increasing inequality; however, it is also possible that strict regulations hinder schools from tailoring inputs to their unique circumstances and serving students to the best of their ability. State regulations are not the result of careful optimization of student achievement but are instead the result of messy political processes (Elmore & Fuhrman, 1995). Thus, the resulting combination of regulations may be inefficient, cumbersome, and a barrier to district improvement.

Empirical evidence of the impact of regulatory flexibility on students has largely come from the charter literature, but charter schools serve a relatively small number of schools and operate within a different competitive and political context than traditional public schools (Education Commission of the States, 2018; National Center for Education Statistics, 2019). Thus, it is unclear the extent to which we should interpret the impact of charter schools as the result of regulatory freedom alone, nor the extent to which we should generalize those results to the population of traditional public schools. There is also a long line of research assessing the impact of *individual* input regulations on student achievement (Boyd et al., 2006; Hanushek, 1986; Kane et al., 2008). This research finds that while some readily measurable school-level inputs matter for student achievement – like class sizes (Krueger, 1999; Lee & Loeb, 2000) – other easily regulated inputs, like teachers' advanced degrees do not create observable shifts in student outcomes (Clotfelter et al., 2007, 2010). This research is helpful for determining which inputs state should consider regulating but does not answer the larger question of whether states should regulate educational inputs *at all*. This paper answers the question: Is the sum total of state input regulations a net positive or net negative for student achievement? By assessing the impact of widespread deregulation in Texas, I provide new empirical evidence of the role of state regulations in producing educational outcomes for a diverse and generalizable population of districts and students.

Prior to 2015, traditional public school districts in Texas were subject to regulations typical of most districts across the United States. Among other requirements, they had to comply with teacher certification standards, class size limits, calendar restrictions, and limitations on staff contracts. But in 2015, this traditional approach to regulation was upended. Today, any Texas district with an acceptable academic and financial rating (a standard met by over 95% of traditional public school districts) can declare *District of Innovation* status and opt out of any regulation which does not apply to the state's charter schools. Using a novel dataset of scraped implementation data posted on school district websites, I document widespread take-up of regulatory exemptions. As of June 2020, nearly 85% of Texas districts have claimed Innovation status, exempting eight regulations on average, including teacher certification requirements, maximum class sizes and minimum instruction time.

Yet, despite enthusiastic take-up, I find that regulatory freedom under the District of Innovation statute caused, at most, a very limited change in observable school district decisions and student outcomes. Using an adapted generalized difference-in-differences strategy (Callaway & Sant'Anna, 2020), I estimate effects on student achievement that are small, negative, and not statistically significant. In math, the average impact was -0.05 school-level standard deviations (SDs). In reading, the average impact was -0.03 SDs. Effects are most negative in the third year of implementation for a small group of first implementers but never reach the point of statistical significance. There was also no significant impact of regulatory freedom on the percent of uncertified teachers, the percent of out-of-field teachers in hard-to-staff subjects, nor on average elementary class sizes. Further, the exemptions claimed by districts do not appear to be a meaningful source of heterogeneous effects, suggesting that no individual regulation lifted by the Texas statute substantially impacted student outcomes.

Taken together, these results demonstrate the stickiness of the status quo in education and suggests that, given stable state accountability statutes and school financing, input regulations have only limited power as a policy-lever. In particular, these results provide strong evidence against the hypothesis that state regulations hinder schools from optimizing student learning. If regulations were a barrier to improvement, then their removal should have increased student outcomes. Yet, I do not find any meaningful changes in either inputs or outcomes as a result of massive deregulation. In Texas, the sum total of state input regulations had at most a minimal influence on student outcomes and school-level decisions. The protective value of state regulations then is either so small as to be statistically indetectable despite the large population or will take more than three years to come into effect. This suggests that state education regulation is a weak policy-lever compared to other state-level interventions like school financing and accountability statutes and that policymakers would be wise to pursue alternative interventions for improving educational outcomes.

Background

Theoretical Justifications for and Against Input Regulations

As accountability statutes came to the forefront of education policy, many imagined them as an alternative to the more traditional input-focused approach to state policy where policymakers govern *how* schools produce student outcomes. They argued that if states were to hold schools and districts accountable for student performance, then they also should remove unproductive regulatory barriers which may make it more difficult for districts to meet their goals (Hanushek, 1996). This view is at least partially bolstered by evidence. When it comes to easily regulated educational inputs, research has shown that “more is not necessarily better” (Cohen et al., 2017, p. 204). Further, even when we suspect that an input is important for student learning, it is unclear at what threshold inputs should be set and it is unlikely that the same threshold applies across all contexts. For example, the now-famous Tennessee STAR experiment found that reducing class sizes to less than 17 students increased K-3 student performance on standardized exams by four percentile points (Krueger, 1999). However, when California implemented a state-wide initiative that reduced K-3 class sizes to approximately 20 students, the program increased the rate of uncertified and inexperienced teachers, demonstrating one potential unintended consequence of using regulation to bring a promising intervention to scale (Jepsen & Rivkin, 2013).

Given the difficulty of creating optimal policy, advocates of deregulation argue that students are better served when their education is designed by people within their community, rather than by distant policymakers and bureaucrats. They argue that school board members, superintendents, principals, and teachers all have greater familiarity with local resources, needs, preferences, priorities, and ideologies than politicians in state capitals (Fuhrman & Elmore, 1995; Moe, 2003). This preference for local control is heavily engrained in American politics; it was the philosophy behind the creation of school boards in colonial Massachusetts (Land, 2002) and the root of small government movements today (Fuhrman, 1993).

Further, even if centralized regulation garners support, there is also the question of whether state officials have the capacity and motivation to set meaningful input policies. Elected officials represent a variety of powerful education constituents, including school superintendents, school boards, teachers, unions and professional associations, and, so, when they do regulate local districts, they often attempt to appease these constituents by setting mandates low enough that most districts exceed them with little effort (Elmore & Fuhrman, 1995). Even when politicians manage to pass meaningful policies, states rarely have centralized education offices with the ability to inspect, monitor, and enforce regulatory compliance. Without the ability to inspect schools and districts in depth, agencies instead rely on districts and schools to honestly document compliance on paper, but, for the most part, districts know that if a regulation is too burdensome, they can ignore it with few repercussions (Elmore & Fuhrman, 1995). Given these challenges in setting and enforcing regulations, many argue they do little more than increase bureaucratic costs.

On the other hand, many consider the goal behind input regulations to be noble; when adequately enforced, education regulations are intended to assure minimum standards throughout the state regardless of demographics, wealth, size, or capacity. Given persistent inequalities in resources, teacher quality, and programmatic offerings (Darling-Hammond & Berry, 2006; Lankford et al., 2002), it is difficult to argue for the removal of minimum quality standards.

Troubled and corrupt districts are of particular concern. Indeed, even though only a few corrupt districts may exist within a state, many state mandates are made with these districts in mind (Fuhrman & Elmore, 1995). Further, input regulations may protect against inappropriately designed incentive systems. Though research demonstrates that accountability policies commonly increase average student performance in tested subjects (Carnoy & Loeb, 2002; T. Dee et al., 2010; Figlio & Loeb, 2011), they can also cause schools to shift resources away from non-tested content and from supports for students on either end of the achievement distribution (T. S. Dee & Jacob, 2011). Input regulations, then, can protect against extreme redistribution of resources. They ensure that even non-tested subjects and students who are unlikely to pass state exams are taught by a certified teacher making at least some minimum salary in a classroom with a maximum number of students at a time.

Previous Experiments with Deregulation

The most common contemporary experiment with deregulation comes in the form of charter schools. Charter schools are publicly funded but privately managed (by non-profit or for-profit operators rather than the public school district). Importantly, state legislation often holds charter schools accountable for student outcomes while providing blanket relief from most regulations other than those related to health, safety, and civil rights (Cohodes, 2018). The impact of charter schools has been extensively researched, demonstrating that their effectiveness is highly variable (Chabrier et al., 2016). In a nationwide study of 32 oversubscribed charter middle schools (which granted admission through random lotteries), Gleason et al. (2010) found that while charters are no more or less effective than traditional public schools at improving student achievement, schools serving the lowest proportions of disadvantaged students had a significant negative impact on student achievement (-0.24 student-level SDs in math) while schools serving high proportions of disadvantaged students had a significant positive impact on student achievement (0.18 SDs in math).

Importantly, researchers have found that many of the practices of highly effective charter are unrelated to regulatory freedom. In a study of New York City charter schools, for example, charter effectiveness was associated with a set of five practices: frequent teacher feedback, data-driven instruction, intensive tutoring, increased instruction time, and high expectations (Dobbie & Fryer, 2013). In contrast, the researchers found no association between commonly regulated inputs – like class size, per pupil expenditures and teacher training – and charter effectiveness. This suggests that charter effectiveness is due, at least in part, to practices that are less common in traditional public school districts, rather than practices that are not possible without deregulation. Similarly, in their study of turnaround schools in Lawrence, Schueler, Goodman, and Deming attribute much of the positive impact of state takeover to intensive small-group instruction over summer breaks, a program which did not require any regulatory changes but may have required the impetus of turnaround as motivation to be bureaucratically feasible (Schueler et al., 2017).

It is important to note two key differences between charter schools and Districts of Innovation. First, Districts of Innovation do not face the same competition for students as charter schools, a key argument underlying the charter school theory of change (Lubienski, 2003). Second, Districts of Innovation face greater political constraints including the presence of traditional hierarchies, bureaucracies, local politics, and school boards. In this sense, Districts of Innovation are most similar, not to charter schools, but to a much smaller experiment with autonomy: Boston pilot schools. Boston's pilot schools have the same regulatory independence as the city's charter schools but remain within the Boston Public School district and so are beholden to district policies and union contracts. In a study comparing the relative effectiveness of Boston charter and public schools, Abdulkadiroğlu and colleagues find large significant gains for charter school students and small negative effects for pilot school students (2011). This finding demonstrates that in the Boston context, regulatory autonomy alone did not result in academic gains for students.

The Texas Context

The biggest difference between the Texas District of Innovation statute and Boston's pilot school program is scale. As of June 2020, 864 districts have declared District of Innovation status. The process of claiming District of Innovation status involves five simple steps wherein the school board: 1) resolves to become a District of Innovation; 2) appoints an Innovation Committee to draft a plan identifying chosen exemptions from the education code; 3) votes on the final plan, 4) posts the Innovation plan on the district website, and 5) notifies the Education Commissioner of their new status. Districts of Innovation are exempt from as many regulations as they choose. The term length of the designation as a District of Innovation cannot exceed 5 years but may be renewed so long as the district continues to meet academic and financial accountability standards.

The Texas District of Innovation statute passed in 2015 and was followed by a steady but enthusiastic response from districts. By the 2016-17 school year, 178 districts had claimed District of Innovation status. An additional 509 districts followed suit in the 2017-18 school year, in addition to 119 districts in 2018-19. By June 2020, 864 traditional school districts (of the 1022 in the state) claimed District of Innovation status. Table 1 provides descriptive characteristics of Districts of Innovation. Compared to the minority of districts who do not adopt this initiative, Districts of Innovation serve a student body that is less Hispanic, whiter, and less economically disadvantaged. However, the geography of Districts of Innovation is representative of the geography of the state; approximately 5% are urban, 24% are suburban, 27% are towns, and 44% are rural. Today, Districts of Innovation are majority district type – encompassing nearly 85% of all Texas school districts.

Data and Measures

This evaluation relies on key information found in district Innovation plans: the regulations from which each district will be exempt and the time period during which the Innovation plan will be active. Like many other online policy documents, Innovation plans contain rich data, but they are found in disparate locations, are stored using diverse media, and require the capacity to turn natural language into structured data in order to extract value. To address these challenges, I use a novel text extraction method described and validated in Author (2019) in which I (1) build a web crawler to visit every district website and download potentially relevant documents; (2) train a convolutional neural network tuned to text classification to identify Innovation plans and discard irrelevant documents; (3) parse each document into a list of regulatory exemptions with the date of implementation.

I link my scraped data to a publicly-available administrative dataset of school-level data from the 2011-12 to 2018-19 school years. The primary outcome of interest in this paper is test scores from the State of Texas Assessments of Academic Readiness (STAAR), a set of standardized exams first implemented in the 2011-12 school year. In particular, I use mathematics scores and reading scores in 3rd through 8th grade as well as English I and Algebra scores in high school to measure performance. I standardize test scores within grade and subject using 2015 means and standard deviations in order to compare results across exams and interpret the magnitude of effects. Though it is districts who claim Innovation status, I treat schools as the unit of analysis, allowing for exploration of within-district variation in implementation and outcomes. This more micro-level data is subject to collinearity within districts standard errors are adjusted to reflect this collinearity.

In exploratory analyses, I also examine school-level inputs which may be impacted by regulatory exemptions. In particular, I examine changes in teacher certification and elementary class sizes. If student outcomes are not affected by regulatory autonomy, this may either be because the changes made to school-level inputs were not important for student outcomes or because schools did not change their inputs. These exploratory analyses help distinguish between the two potential explanations. I focus on teacher qualifications both because of the popularity of the exemption (certification requirements are exempted by 87% of Districts of Innovation) and because fears regarding uncertified teachers were a popular critique of the statute (Texas Classroom Teachers Association, 2017). To explore changes in teacher certification, I examine the percent of uncertified teachers and the percent of out-of-field teachers in hard-to-staff subjects (secondary math, science, and career and technical education). These are teachers who are certified, but not in the subject they are teaching. I focus on average class sizes in elementary schools again because of the popularity of the elementary class size exemption (44% of school districts) and because of evidence regarding the importance of class sizes in elementary grades (Krueger, 1999).

Finally, I explore the heterogeneous impact of the statute by type of exemptions claimed, district-level geography, and school-level demographics and prior achievement. These characteristics may be important predictors of effects if, as in the charter literature, urban and non-urban districts implement different practices or if those practices are most effective for some student subgroups. Further, urbanicity and student demographics may be correlated with budgetary, political, and knowledge constraints. If these constraints interact with regulatory freedom, it will be important to ensure that the District of Innovation statute does not exacerbate current inequities.

Empirical Strategy

In my impact analyses, I define the treatment as District of Innovation status. This means that treatment is defined as access to blanket regulatory flexibility on schools within Districts of Innovation. Due to substantive differences between Districts of Innovation and the remaining traditional public school districts (see Table 1), I limit my analytic sample to schools within Districts of Innovation and capitalize on variation in implementation dates. In other words, this strategy facilitates comparisons between treated schools who took up regulatory flexibility at different points in time and does not examine outcomes for untreated schools. One typical approach to assessing impacts with variation in treatment timing is to employ an event study or generalized difference-in-differences design with two-way fixed effects. However, recent research demonstrates that weighting schemes embedded in these approaches produce biased and unintuitive impact estimates when effects vary with the length of exposure to the policy (Callaway & Sant'Anna, 2020; Goodman-Bacon, 2021). I therefore follow the advice of Callaway and Sant'Anna and estimate the impact of District of Innovation status using a series of simple difference-in-difference (DID) estimates.

For each implementation cohort (2017, 2018, and 2019) and each year, I estimate a simple 2-group, 2-time-period DID where the cohort of interest serves as the treatment group and schools within Districts of Innovation that have not yet implemented their Innovation plan serve as the comparison group. I include 2020 implementers in the comparison group for each other cohort as I do not have outcomes for this group due to Covid-19 related test cancellations. For each cohort, the difference between pre-treatment and post-treatment outcomes is estimated using the last pre-treatment year before that that cohort's Innovation plans are active. For example, the effect of District of Innovation status in 2017 for 2017 implementers is estimated using 2016 as the pre-treatment year and with 2018, 2019, and 2020 implementers as the comparison group. The effect of District of Innovation status for 2017 implementers after three years of implementation (in 2019) is also estimated using 2016 as the pre-treatment year, but with only 2020 implementers in the comparison group.

Using the last pre-treatment year limits the likelihood of confounding events that occur within the implementation group that do not occur within the comparison group, but it does not address potential anticipatory effects of adoption (i.e., if school leaders change their decisions because they know that their school will soon be within a District of Innovation). In this case, however, no anticipatory effects are detected (see the placebo effect estimates within the last pre-treatment year in Tables 5 and 6).

Formally, I estimate,

$$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 TreatPost_{st} + \varepsilon_{st}$$

where Y_{st} is the outcome of interest for school s in year t. $Treat_s$ is an indicator for whether the school is within the cohort of interest (for example, whether the school is a 2017 implementer), $Post_t$ is an indicator for whether the year is post-treatment for the cohort (in the example, post-2017), and the coefficient on $TreatPost_{st}$ is the estimated impact of District of Innovation status for the cohort and year. In total, I estimate six DID treatment effect estimates: the impact of District of Innovation status for the 2017 cohort in 2017, 2018, and 2019, the impact for the 2018 cohort in 2018 and 2019, and the impact for the 2019 cohort in 2019. I also estimate 16 pre-treatment effect estimates which serve as placebo tests for pre-trend assumptions. In addition to estimating individual year by cohort effects, I aggregate the estimates using two schemes: 1) an average treatment effect estimate across all cohorts and times weighted by the size of the cohort; and 2) dynamic average treatment effect estimates after one, two, and three years of implementation.

Threats to Validity

The key assumption in the DID strategy is that each cohort would have experienced the same change in outcomes as the comparison group, if not for changes resulting from District of Innovation status itself. This is colloquially known as the parallel trends assumption. An example threat to this assumption would occur if early adopters were eager to claim Innovation status because they were on a different student achievement trajectory than their peers. This threat is not particularly likely in the District of Innovation case because, unlike many other educational interventions, the District of Innovation statute is neither meant to be a reform strategy for struggling districts nor an incentive for successful districts. Instead, it was borne out of an argument that traditional public districts should have the same autonomy from regulation as the state's charters.

Nonetheless, I test the exogeneity of treatment timing using two approaches. First, I estimate the "impact" of District of Innovation status on school-level reading and math achievement in the years before the Innovation plan was implemented. This is a test of parallel trends in the pre-treatment period. If trends were not parallel between groups before implementation, it would be less plausible that they would have been parallel if not for the effects of District of Innovation status. The unhighlighted rows of Tables 5 and 6 show these tests of parallel trends for reading and math, respectively. Across the two tables, the DID model fails to identify a treatment effect for any of the pre-treatment year by cohort combinations.

Second, I estimate the impact of District of Innovation status on the proportion of students that are Black, Hispanic, on free or reduced price lunch (FRPL), and who have an individualized education plan. In the post-treatment years, this assesses whether claiming District of Innovation status causes students to move from one school to another. In the pre-treatment years, this provides evidence on whether demographic changes vary by cohort, which would indicate that cohorts are subject to different state-wide trends. Appendix A shows the results for each of these demographic variables; there is no estimated impact of District of Innovation status for any demographic variable or any time period.

Results

Exemptions Claimed by Districts of Innovation

Table 2 presents the fifteen most commonly exempted regulations and the proportion of Districts of Innovation exempting that statute. On average, Districts of Innovation claim eight exemptions spanning a range of educational inputs. The most popular exemptions concern school schedules, teacher certification, class sizes, teacher contracts, and student behavior. By far, the most popular exemption is the requirement that school districts not begin instruction before the fourth Monday in August; 98% of Districts of Innovation have exempted this requirement, allowing them to start the school year earlier. On the other hand, other popular scheduling exemptions give districts the freedom to shorten the school year. For example, 28% of districts have exempted the requirement that the last day of school not occur before May 15th and 42% have exempted the requirement that schools operate for at least 75,600 minutes each year (an average of seven hours a day if the school is open for 180 days).

The second most popular category of exemptions concerns teacher certification and teaching conditions. Notably, 87% of Districts of Innovation are no longer required to hire certified teachers or to ensure that teachers are certified in the field they teach. Further, just under half of all Districts of Innovation have exempted the requirement that elementary class sizes do not exceed 22 students and that teachers be fired or tenured within three years. Additionally, 36% of Districts of Innovation have exempted the requirement that teacher contracts extend for at least 187 days a year and 20% have exempted the requirement that teachers are evaluated at least once a year using performance criteria developed by the education commissioner or by the district themselves.

A final category of popular exemptions generally pertains to school responses to student behavior. The most popular exemption in this category, with 26% of Districts of Innovation exempting, is the regulation that requires students to attend nine-tenths of a course in order to receive credit. A fifth of districts have also exempted the requirement that schools designate a single person as the campus behavior coordinator who handle behavior referrals and a fifth have exempted requirements regarding the acceptance of student transfers.

Patterns in Exemptions

Table 3 disaggregates the percent of districts exempting each regulation by urbanicity. Table B1 provides additional descriptive statistics of teacher and student characteristics for exempting Districts of Innovation compared to non-exempting Districts of Innovation. Table 3 demonstrates some categories of exemptions are universally popular, like early school start dates, while others show a clear divide in popularity between urban and rural districts. Compared to urban and suburban districts, rural districts are more likely to exempt regulations related to teacher certification, teacher tenure, and minimum service days required for teachers. For example, while 91% of rural districts have excepted teacher certification requirements, only 76% of urban districts have done the same. Even more striking, while 55% of rural districts have exempted the minimum service days required for teachers, only 3% of urban districts claimed the same exception. These patterns suggest that rural districts struggle to recruit highly effective teachers and so may wish to use regulatory freedom to consider non-traditional uncertified applicants, to entice new applicants by decreasing workdays, and to reduce the number of teachers they are forced to fire or provide with tenure. Conversely, urban districts are less likely to exempt statutes concerning hiring and contracts and more likely to exempt statutes concerning expectations for students. These exemptions allow urban schools to lower attendance standards and to task multiple personnel with responding to student misbehavior.

The Impact of District of Innovation Status on School-Level Academic Achievement

Despite wide ranging exemptions claimed by districts, the impact of District of Innovation status on academic achievement is negligible. In math, the average impact is -0.05 SDs and not statistically significant (see Table 4). Table 5 disaggregates this effect by year and implementation group. In the first group of implementers, I observe small, negative estimates in the first and second years of implementation (-0.04 SDs and -0.05 SDs, respectively) and a larger, not statistically significant effect in the third year of implementation (-0.12 SD). Note that standard errors are substantially larger in the third year (2019) for the 2017 implementers. This is because while the comparison group in 2017 consists of 2018, 2019, and 2020 implementers, the comparison group in 2019 only contains 2020 implementers as every other cohort is now formally accounted for as treated. The pattern of effects for the 2017 cohort is replicated by the remaining implementation cohorts; there is no impact of District of Innovation status on math scores in the first year of implementation for the 2018 or 2019 implementers and a small negative impact of -0.04 SDs in the second year of implementation for 2018 implementers (Table 5). In reading, effects are also slightly negative (-0.03 SDs, on average) and never reach a level of statistical significance (see Table 6). Like with math, the largest effects are in the third year of implementation for the first cohort (-0.07 SDs).

Appendix C shows the average impact estimates for individual tests. After Bonferroni adjustment for multiple hypothesis testing, no subjects show statistically significant impacts, nor does there appear to be any pattern to the magnitude of effects. For example, while the largest effect estimates occur in 7th grade math, other middle school grades (6th and 8th grade) have effect estimates that are negligible. This seems to suggest that any heterogeneous effects by subject are simply due random variation.

Finally, Table 7 explores the potential for heterogeneous subgroup effects by urbanicity, prior student achievement, and student demographics. On the left panel, the table displays impact estimates for rural schools, schools with average STAAR scores in 2016 that are below the state mean, schools with fewer Hispanic students (less than the state median proportion of 0.41), and schools with fewer Black students (less than the state median proportion of 0.06). The right panel displays impact estimates for urban schools, schools with higher average STAAR scores, schools with more Hispanic students, and schools with more Black students. Effects are not significant for any sub-group (nor are the differences between subgroups), but are slightly more negative for urban schools, schools with lower prior achievement, and for schools serving higher proportions of Black students.

Heterogeneous Impacts by District Exemptions

To explore whether some regulatory exemptions are more important for student outcomes than others, I explore heterogeneous impacts by the type of exemptions claimed by districts. Tables 8 and 9 present sub-group impact estimates for exempters and non-exempters for nine types of exemptions. Across all exemptions, only one difference in subgroup effects is statistically significant. While schools within districts that exempt the earliest first day of instruction experience an impact of -0.04 SDs on math and -0.02 SDs on reading, those that do *not* exempt the earliest first day experience an impact of -0.48 SDs and -0.40 SDs for math and reading, respectively. However, this effect is for a very small subset of schools within just 14 districts that *do not* exempt a regulation. Thus, this sub-group effect likely has little to do with the regulation itself. There is no reasonable theory for how the act of choosing not to exempt a regulation would cause test scores to decrease. Instead, this sub-group effect likely relates to the *kind of schools* that are within districts that do not want to use their District of Innovation status to begin the school year earlier. These schools tend to have much lower school achievement than the average District of Innovation and tend to serve a more Hispanic and economically disadvantaged population (Appendix B). However, given the extremity of these results and the small population of schools within this sub-group, generalizations from this sub-group results may be unwise.

Impact of School-Level Inputs

Finally, Table 10 presents the average impact estimates on five school level inputs: the proportion of uncertified teachers, the proportion of secondary math, science, and career and technical teachers that are out of field (certified, but not in that subject), and average elementary class sizes. None of the impact estimates are significant or substantial. Despite substantial take-up of certification and class size exemptions, there do not appear to be any changes at the school level. District of Innovation status does not result in an increase in uncertified teachers, in out-of-field teachers in math or science, or in class sizes. The absence of school-level changes for these inputs may provide some explanation for the null effects on student achievement.

Discussion

There are two overarching hypotheses about the role of regulations in improving public education. In one camp, experts argue that regulations hinder schools from tailoring inputs to student needs and from serving students to the best of their ability (Fuhrman & Elmore, 1995; Hanushek, 2003). In another camp, experts argue that regulations ensure some minimum standard of quality and protect students against negative outcomes (Darling-Hammond & Berry, 2006; Fuhrman & Elmore, 1995). This paper demonstrates that within the context of Districts of Innovation, neither hypothesis holds much sway. Though the District of Innovation statute has resulted in over 5,000 regulatory exemptions in staffing, school time, human resource policies, and student behavior policies, there is neither substantial evidence of a meaningful impact on student achievement, nor is there a detectable impact on intermediate outcomes like the hiring of uncertified teachers or elementary class sizes.

If regulations impeded schools from reaching their full academic potential, then deregulation under the Texas District of Innovation statute would cause a positive impact on school achievement. However, across outcomes, years, and subgroups, the impact of the statute is never positive. On the other hand, regulations also do not appear to offer substantial protections for students against negative outcomes. Though the effects of regulatory autonomy in Texas are negative, they are small and do not reach a level of statistical significance. On average, the impact on achievement in mathematics is -0.05 SDs and in reading it is -0.03 SDs.

Disaggregating the average effect by year and implementation group shows that the largest impact is in the third year of implementation; the earliest adopters of the District of Innovation statute experienced a negative impact -0.12 SDs on scores in mathematics after three years of implementation. However, this three-year effect requires an important caveat. First, because this paper uses school-level data, the observed effect sizes cannot be directly compared to other interventions estimating effects using student-level data. While a -0.12 SD effect may be considered moderate when considering student-level test scores, it is less substantial when considering school-level test scores. As a quick back of the envelope translation, the school-level SD in Texas is about twice as large as a student-level SD, so we can roughly divide District of Innovation effects by two to better compare it to other interventions. The third-year effect in mathematics, then, is roughly -0.06 student-level SDs. Thus, any protective value of regulations is small and statistically indetectable within three years of implementation.

However, this finding has three key limitations. First, this study does not answer questions regarding the efficacy of individual regulations; it only assesses the impact of lifting the full set of input regulations passed by the Texas legislature. Second, the study only estimates the impact of the statute on a subset of outcomes (math and reading), inputs (certification and elementary class sizes), and years (up to three years of implementation for early implementers). This approach does not allow me to estimate potential effects of exemptions that only result in long-term changes. For example, I am not able to observe whether a teacher has been granted tenure or remains on a probationary contract and the impact of lifting tenure requirements may not result in changes in student outcomes for several years. Changes associated with this exemption, therefore, may be undetected in this analysis. Third, we may be most concerned about the impact of deregulation on outcomes for which schools are not held accountable. Changes in student experiences in untested subjects and student's socio-emotional learning, for example, are not accounted for in the current measures. This paper therefore provides no evidence on those outcomes.

Despite these limitations, this paper makes three important contributions. First, the paper demonstrates the potential use of web scraping and natural language processing methods in policy evaluation. The identification strategy employed here, comparing adoption cohorts to one-another, leveraged data which were semi-automatically scraped from publicly-posted District of Innovation plans. Many implementation details in education policies are documented in similar forms, and therefore this technique has the potential to facilitate similar implementation analyses at scale. Second, this paper is the first to estimate the impact of education deregulation outside the context of charter schools for a generalizable population of traditional public school districts. While previous studies have examined the impact of deregulation for a select subset of schools (Abdulkadiroğlu et al., 2011), the District of Innovation statute provides the opportunity to study the impact of deregulation at scale. Third, by analyzing the impact of the statute on key educational inputs like teacher certification and class sizes, this paper documents the limited impact of regulatory flexibility on school-level decisions. Many districts in this context articulate the desire for certain regulatory freedoms by explicitly naming exempted inputs (e.g., teacher certification), but school leaders largely do not act on these freedoms (e.g., by hiring uncertified teachers). This demonstrates the stubbornness of the status quo in education and the hesitancy of school officials to experiment, even without regulatory barriers (Century & Cassata, 2016; Lubienski, 2003). Taken as a whole, this study suggests that, given stable accountability statutes and financing, state government regulation of educational inputs is not significant driver of student achievement in either direction.

Bibliography

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots. *Quarterly Journal of Economics*, *126*(2), 699–748. https://doi.org/10.1093/qje/qjr017

Anglin, K. L. (2019). Gather-Narrow-Extract: A Framework for Studying Local Policy Variation Using Web-Scraping and Natural Language Processing. *Journal of Research on Educational Effectiveness*, *12*(4), 685–706. https://doi.org/10.1080/19345747.2019.1654576

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement. *Education Finance and Policy*, *1*(2), 176–216. https://doi.org/10.1162/edfp.2006.1.2.176

Callaway, B., & Sant'Anna, P. H. C. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.12.001

Carnoy, M., & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331. https://doi.org/10.3102/01623737024004305

Century, J., & Cassata, A. (2016). Implementation Research: Finding Common Ground on What, How, Why, Where, and Who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732x16665332

Chabrier, J., Cohodes, S., & Oreopoulos, P. (2016). What Can We Learn from Charter School Lotteries? *Journal of Economic Perspectives*, *30*(3), 57–84. https://doi.org/10.1257/jep.30.3.57

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and Why Do Teacher Credentials Matter for Student Achievement?* (No. 12828; NBER Working Paper Series, pp. 1–54). NBER. https://www.nber.org/system/files/working_papers/w12828/w12828.pdf

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. *Journal of Human Resources*, *45*(3), 655–681. https://doi.org/10.1353/jhr.2010.0023

Cohen, D. K., Spillane, J. P., & Peurach, D. J. (2017). The Dilemmas of Educational Reform. *Educational Researcher*, *47*(3), 0013189X1774348. https://doi.org/10.3102/0013189X17743488

Cohodes, S. (2018). Charter Schools and the Achievement Gap. *Future of Children*, *Policy*, 1–16.

Darling-Hammond, L., & Berry, B. (2006). Highly Qualified Teachers for All. *Educational Leadership*, *64*(3), 14.

Dee, T., Jacob, B., Hoxby, C., & Ladd, H. F. (2010). The Impact of No Child Left Behind on Students, Teachers, and Schools [with Comments and Discussion]. *Brookings Papers on Economic Activity*, 149–207.

Dee, T. S., & Jacob, B. (2011). The Impact of No Child Left Behind on Students, Teachers, and Schools. *Journal of Policy Analysis and Management*, *30*(3), 418–446. https://doi.org/10.1002/pam.20586

Dobbie, W., & Fryer, R. G. (2013). Getting Beneath the Veil of Effective Schools: Evidence From New York City. *American Economic Journal: Applied Economics*, *5*(4), 28–60. https://doi.org/10.1257/app.5.4.28

Education Commission of the States. (2005). *State Class-size Reduction Measures*.

Education Commission of the States. (2017). *Teacher License Reciprocity: All data points for all states*.

Education Commission of the States. (2018). *50 State Comparison: Charter School Policies*.

Education Commission of the States. (2019). *50-State Comparison: Minimum Number of Days or Hours per School Year*.

Elmore, R. F., & Fuhrman, S. (1995). *Opportunity to Learn and the State Role in Education*. National Governors Association. https://ncela.ed.gov/files/rcd/BE022357/Opportunity_to_Learn.pdf

Figlio, D., & Loeb, S. (2011). School Accountability. In *Handbook of the Economics of Education* (1st ed., Vol. 3, pp. 343–421). Elsevier B.V. https://doi.org/10.1016/B978-0-444-53429-3.00008-9

Fuhrman, S. H. (1993). The Politics of Coherence. In *Designing Coherent Education Policy: Improving the System* (pp. 1–34).

Fuhrman, S. H., & Elmore, R. F. (1995). Ruling Out Rules: The Evolution of Deregulation in State Education Policy. *Consortium for Policy Research in Education*, *708*(2).

Gleason, P., Clark, M., Clark-Tuttle, C., & Dwoyer, E. (2010). *The Evaluation of Charter School Impacts: Executive Summary* (NCEE 2010-4029; pp. 1–14). http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED510573

Goodman-Bacon, A. (2021). Difference-in-Difference with Variation in Treatment Timing. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2021.03.014

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, *24*(3), 1141–1177. https://doi.org/10.2307/2725865

Hanushek, E. A. (1996). Outcomes, Costs, and Incentives in Schools. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's Schools: The Role of Incentives* (pp. 29–52).

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies. *The Economic Journal*, *113*(February), F64–F98.

Jepsen, C., & Rivkin, S. G. (2013). Class Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size. *Journal of Human Resources*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Kane, T., Rockoff, J., & Staiger, D. (2008). What Does Certification Tell s About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631. https://doi.org/10.1016/j.econedurev.2007.05.005

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, *114*(2), 497–532.

Land, D. (2002). Local School Boards Under Review: Their Role and Effectiveness in Relation to Students' Academic Achievement. *Review of Educational Research*, *72*(2), 229–278. https://doi.org/10.3102/00346543072002229

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62. https://doi.org/10.3102/01623737024001037

Lee, V. E., & Loeb, S. (2000). School Size in Chicago Elementary Schools: Effects on Teachers' Attitudes and Students' Achievement. *American Educational Research Journal*, *37*(1), 3–31. https://doi.org/10.3102/00028312037001003

Lubienski, C. (2003). Innovation in Education Markets: Theory and Evidence on the Impact of Competition and Choice in Charter Schools. *American Educational Research Journal*, *40*(2), 395–443. https://doi.org/10.3102/00028312040002395

Moe, T. M. (2003). Politics, Control and the Future of School Accountability. In *No Child Left Behind? The Politics and Practice of School Accountability* (pp. 80–106).

National Center for Education Statistics. (2019). *Public Charter School Enrollment*.

Schueler, B. E., Goodman, J. S., & Deming, D. J. (2017). Can States Take Over and Turn Around School Districts? Evidence From Lawrence, Massachusetts. *Educational Evaluation and Policy Analysis*, *39*(2), 311–332. https://doi.org/10.3102/0162373716685824

Texas Classroom Teachers Association. (2017). *Districts of Innovation: The Reality*. Texas Classroom Teachers Association. https://tcta.org/node/14394-districts_of_innovation_the_reality

Table 1

2015-16 Characteristics of Eligible Traditional Public School Districts vs. Districts of Innovation

	Traditional Public School District Average *N = 157*	District of Innovation Difference *N = 864*
District Characteristics		
Urban	0.08	-0.03
		(0.02)
Suburban	0.2	0.04
		(0.04)
Town	0.25	0.02
		(0.04)
Rural	0.47	-0.03
		(0.04)
Teacher Characteristics		
Experience Teaching	11.91	0.53*
		(0.22)
Experience in District	6.98	0.18
		(0.17)
Percent Teacher Turning Over	21.57	-2.74***
		(0.84)
Student Teacher Ratio	12.57	0.15
		(0.21)
Student Characteristics		
% Hispanic	0.54	-0.18***
		(0.02)
% White	0.37	0.16***
		(0.02)
% Black	0.06	0.01
		(0.01)
% Free/Reduced Lunch	0.66	-0.1***
		(0.02)
Average STAAR Performance	-0.25	0.34***
		(0.06)

Note. Column 1 presents the mean baseline characteristics of traditional public school districts that were eligible to declare District of Innovation status. Column 2 presents the mean difference for Districts of Innovation. Standard errors are reported in parentheses. *** = p<.001, ** = p<.01, * = p<.05

Table 2

Proportion Districts of Innovation Exempting the Most Common Regulations

Regulation	Proportion
School Schedules	
25.081 – Earliest First Day of Instruction	0.98
25.081 – Minimum Minutes of Operation	0.42
25.0812 – Earliest Last Day of Instruction	0.28
25.082 – Pledge of Allegiance and Minute of Silence	0.25
Certification	
21.003 – Teacher Certification Required	0.87
21.057 – Notice of Uncertified Teacher	0.33
21.053 – Presentation of Teacher Certificates	0.29
Professional Contracts	
21.102 – Maximum Probationary Contract Length	0.52
21.401 – Minimum Service Days Required for Teachers	0.36
21.352 – Teacher Evaluation Requirements	0.20
Class Size	
25.112 – Elementary Class Size Maximum	0.44
25.113 – Notice of Class Size	0.37
Responses to Students	
25.092 – Minimum Attendance for Class Credit	0.26
37.0012 – Campus Behavior Coordinator	0.22
25.036 – Student Transfer Requirements	0.24

Note. Statistics are as of June 2020. Data were scraped from District of Innovation plans posted on school district websites. The table presents the top fifteen most exempted regulations.

Table 3

Proportion of Districts of Innovation Exempting Regulations by Urbanicity

Description	Count	Urban	Suburban	Town	Rural	F-Test P-Value
School Schedules						
25.0811 - Earliest First Day of Instruction	849	1	0.98	0.98	0.98	0.86
25.081 - Minimum Minutes of Operation	**360**	**0.5**	**0.38**	**0.36**	**0.46**	**0.05**
25.0812 - Earliest Last Day of Instruction	245	0.24	0.24	0.28	0.32	0.26
25.082 - Pledge of Allegiance and Minute of Silence	215	0.24	0.2	0.27	0.26	0.31
Class Size						
25.112 - Class Size Maximum	379	0.39	0.43	0.45	0.44	0.91
25.113 - Notice of Class Size	315	0.37	0.36	0.39	0.36	0.89
Teacher Certification						
21.003 - Teacher Certification Required	**755**	**0.76**	**0.82**	**0.88**	**0.91**	**<.01**
21.053 - Presentation of Teacher Certificates	252	0.32	0.24	0.32	0.3	0.33
21.057 - Notice of Uncertified Teacher	281	0.32	0.25	0.35	0.35	0.08
Professional Contracts						
21.102 - Maximum Probationary Contract Length	**452**	**0.21**	**0.42**	**0.59**	**0.58**	**<.01**
21.401 - Minimum Service Days Required for Teachers	**308**	**0.03**	**0.14**	**0.29**	**0.55**	**<.01**
21.352 - Teacher Evaluation Requirements	172	0.24	0.21	0.2	0.19	0.9
Student Behavior						
25.092 - Minimum Attendance for Class Credit	**228**	**0.53**	**0.36**	**0.27**	**0.18**	**<.01**
37.0012 - Designation of Campus Behavior Coordinator	**190**	**0.37**	**0.28**	**0.27**	**0.14**	**<.01**
25.036 - Transfers	**207**	**0.03**	**0.09**	**0.26**	**0.33**	**<.01**

Note. Exemption statistics are as of June 2020. Regulations with significant differences ($p<0.05$) between urbanicity groups are bolded.

Table 4

Aggregated Estimates of the Impact of District of Innovation Status on School-Level Academic Achievement

Math	Overall	Dynamic
	-0.05	**t=1**
	(0.03)	-0.02
		(0.02)
		t=2
		-0.05
		(0.03)
		t=3
		-0.12
		(0.07)
Reading	Overall	Dynamic
	-0.03	**t=1**
	(0.02)	-0.01
		(0.01)
		t=2
		-0.03
		(0.03)
		t=3
		-0.07
		(0.05)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. Overall estimates are a weighted average of all post-treatment years and implementation cohorts. 2019 implementers have one post-treatment outcome year, 2018 implementers have two post-treatment years, and 2017 implementers have three post-treatment years. 2020 implementers are included in the comparison group for all implementation cohorts. Standard errors are bootstrapped. No treatment effect estimates are significant at 5% significance level. *p<0.05

Table 5

Estimates of the Impact of District of Innovation Status on Math Disaggregated by Implementation Cohort and Time

	2017 Implementers	2018 Implementers	2019 Implementers
2013	0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.03)
2014	0.0	-0.0	-0.0
	(0.02)	(0.02)	(0.02)
2015	-0.04	0.02	0.02
	(0.02)	(0.02)	(0.02)
2016	0.0	0.0	0.01
	(0.02)	(0.02)	(0.02)
2017	-0.04	0.02	-0.03
	(0.02)	(0.03)	(0.03)
2018	-0.05	-0.01	-0.02
	(0.04)	(0.02)	(0.03)
2019	-0.12	-0.04	-0.0
	(0.07)	(0.04)	(0.03)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. The rows highlighted in gray indicate a treatment effect estimate. The remaining rows are placebo tests of an effect before implementation. 2020 implementers are included in the comparison group for all cohorts. Standard errors are in parentheses and bootstrapped. No treatment effect estimates are significant at 5% significance level. *p<0.05

Table 6

Estimates of the Impact of District of Innovation Status on Reading Disaggregated by Implementation Cohort and Time

	2017 Implementers	2018 Implementers	2019 Implementers
2013	-0.0	-0.0	0.01
	(0.01)	(0.01)	(0.02)
2014	-0.0	-0.01	0.01
	(0.02)	(0.01)	(0.01)
2015	0.03	-0.03	0.01
	(0.02)	(0.02)	(0.02)
2016	-0.01	0.02	-0.02
	(0.02)	(0.02)	(0.02)
2017	-0.0	0.01	-0.02
	(0.02)	(0.02)	(0.02)
2018	-0.01	-0.02	-0.02
	(0.03)	(0.02)	(0.02)
2019	-0.07	-0.04	0.0
	(0.04)	(0.04)	(0.03)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. The rows highlighted in gray indicate a treatment effect estimate. The remaining rows are placebo tests of an effect before implementation. 2020 implementers are included in the comparison group for all cohorts. Standard errors are in parentheses and bootstrapped. No treatment effect estimates are significant at 5% significance level. *p<0.05

Table 7

Estimates of the Impact of District of Innovation Status, Disaggregated by Demographic Subgroup

	Below State Median	Above State Median
Math		
Rural/Urban	-0.03	-0.08
	(0.06)	(0.04)
Prior Achievement	-0.07	-0.0
	(0.04)	(0.03)
Percent Hispanic Students	-0.04	-0.05
(Median = 41%)	(0.03)	(0.04)
Percent Black Students	-0.03	-0.07
(Median = 6%)	(0.03)	(0.03)
Reading		
Rural/Urban	-0.03	-0.01
	(0.04)	(0.03)
Prior Achievement	-0.03	-0.01
	(0.03)	(0.02)
Percent Hispanic Students	-0.02	-0.03
(Median = 41%)	(0.02)	(0.03)
Percent Black Students	-0.03	-0.01
(Median = 6%)	(0.03)	(0.02)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. Estimates are a weighted average of the treatment effect for all post-treatment years and implementation groups. 2019 implementers have one post-treatment outcome year, 2018 implementers have two post-treatment years, and 2017 implementers have three post-treatment years. 2020 implementers are included in the control group for all implementation cohorts. Except for urbanicity, schools are grouped by whether they were below or above the median for the school characteristic. Standard errors are bootstrapped. For urbanicity, subgroups represent rural school and urban schools (left to right). After Bonferroni adjustment, no treatment effect estimates are significant at 5% significance level, nor are the differences between sub-groups. *p<0.05

Table 8

Estimates of the Impact of District of Innovation Status on School Math Achievement, Disaggregated by Exemption Status

	Non-Exempters	Exempters
First Day of Instruction	-0.48***	-0.04
	(0.08)	(0.02)
Minimum Minutes of Operation	-0.04	0.01
	(0.03)	(0.04)
Last Day of Instruction	-0.06	0.02
	(0.03)	(0.05)
Certification	-0.11	-0.03
	(0.06)	(0.03)
Probation	-0.03	-0.06
	(0.03)	(0.04)
Service Days	-0.03	-0.12*
	(0.03)	(0.04)
Teacher Evaluation	-0.05	-0.01
	(0.03)	(0.06)
Elementary Class Size	-0.02	-0.07
	(0.03)	(0.04)
Behavior Coordinator	-0.06	-0.01
	(0.03)	(0.04)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. Estimates are a weighted average of the treatment effect for all post-treatment years and implementation groups. 2019 implementers have one post-treatment outcome year, 2018 implementers have two post-treatment years, and 2017 implementers have three post-treatment years. 2020 implementers are included in the control group for all implementation cohorts. Standard errors are bootstrapped. *** = p<.001, ** = p<.01, * = p<.05

Table 9

Estimates of the Impact of District of Innovation Status on School Reading Achievement, Disaggregated by Exemption Status

	Non-Exempters	**Exempters**
First Day of Instruction	-0.40***	-0.02
	(0.07)	(0.02)
Minimum Minutes of Operation	-0.03	0.03
	(0.02)	(0.03)
Last Day of Instruction	-0.04	0.05
	(0.02)	(0.03)
Certification	-0.07	-0.01
	(0.04)	(0.02)
Probation	-0.03	-0.02
	(0.03)	(0.02)
Service Days	-0.03	-0.03
	(0.02)	(0.04)
Teacher Evaluation	-0.03	0.02
	(0.02)	(0.02)
Elementary Class Size	-0.01	-0.05
	(0.03)	(0.02)
Behavior Coordinator	-0.03	-0.00
	(0.02)	(0.03)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. Estimates are a weighted average of all post-treatment years and implementation groups. 2019 implementers have one post-treatment outcome year, 2018 implementers have two post-treatment years, and 2017 implementers have three post-treatment years. 2020 implementers are included in the control group for all implementation cohorts. Standard errors are bootstrapped. *** = p<.001, ** = p<.01, * = p<.05

Table 10

Aggregated Estimates of the Impact of District of Innovation Status on School-Level Inputs

Proportion Uncertified Teachers	-0.0
	(0.0)
Proportion Out-of-Field Secondary Math Teachers	0.0
	(0.02)
Proportion Out-of-Field Secondary Science Teachers	-0.0
	(0.02)
Proportion Out-of-Field Secondary Career and Technical Teachers	0.01
	(0.01)
Average Elementary Class Size	-0.25
	(0.24)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. Overall estimates are a weighted average of all post-treatment years and implementation groups. 2019 implementers have one post-treatment outcome year, 2018 implementers have two post-treatment years, and 2017 implementers have three post-treatment years. 2020 implementers are included in the control group for all implementation cohorts. Standard errors are bootstrapped. No treatment effect estimates are significant at 5% significance level.
*p<0.05

Appendix A

The Impact of District of Innovation Status on Student Demographics

Table A1

Difference-in-Differences Estimates of the Impact of District of Innovation Status on the Percent of Students who are Black, Disaggregated by Cohort and Time

Group	Time	Treatment Effect	Lower 95% CI	Upper 95% CI
2017	2013	0.00	0.00	0.00
2017	2014	0.00	0.00	0.00
2017	2015	0.00	0.00	0.00
2017	2016	0.00	0.00	0.00
2017	2017	0.00	0.00	0.00
2017	2018	0.00	0.00	0.01
2017	2019	0.01	0.00	0.02
2018	2013	0.00	0.00	0.00
2018	2014	0.00	0.00	0.00
2018	2015	0.00	0.00	0.00
2018	2016	0.00	0.00	0.00
2018	2017	0.00	0.00	0.00
2018	2018	0.00	0.00	0.01
2018	2019	0.01	-0.01	0.02
2019	2013	0.00	0.00	0.00
2019	2014	0.00	0.00	0.00
2019	2015	0.00	0.00	0.00
2019	2016	0.00	0.00	0.00
2019	2017	0.00	0.00	0.00
2019	2018	0.00	-0.01	0.02
2019	2019	0.00	0.00	0.01

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. 2020 implementers are included in the comparison group for all implementation cohorts. The rows highlighted in gray indicate a treatment effect estimate. The remaining rows are placebo tests of an effect before implementation. Confidence intervals are estimated with bootstrapped standard errors, adjusting for clustering within districts. No treatment effect estimates are significant at 5% significance level.

Table A2

Estimates of the Impact of District of Innovation Status on the Percent of Students who are Hispanic, Disaggregated Implementation Cohort and Time

Group	Time	Treatment Effect	Lower 95% CI	Upper 95% CI
2017	2013	0.00	-0.01	0.00
2017	2014	0.00	0.00	0.01
2017	2015	0.00	-0.01	0.00
2017	2016	0.00	-0.01	0.00
2017	2017	0.00	0.00	0.00
2017	2018	0.00	-0.01	0.00
2017	2019	-0.01	-0.03	0.01
2018	2013	0.00	0.00	0.01
2018	2014	0.00	-0.01	0.00
2018	2015	0.00	0.00	0.00
2018	2016	0.00	0.00	0.00
2018	2017	0.00	0.00	0.00
2018	2018	0.00	-0.01	0.00
2018	2019	-0.01	-0.03	0.01
2019	2013	0.00	0.00	0.00
2019	2014	0.00	0.00	0.01
2019	2015	0.00	0.00	0.00
2019	2016	0.00	0.00	0.01
2019	2017	0.00	-0.01	0.00
2019	2018	0.00	-0.01	0.01
2019	2019	0.00	-0.01	0.01

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. 2020 implementers are included in the comparison group for all implementation cohorts. The rows highlighted in gray indicate a treatment effect estimate. The remaining rows are placebo tests of an effect before implementation. Confidence intervals are estimated with bootstrapped standard errors, adjusting for clustering within districts. No treatment effect estimates are significant at 5% significance level.

Table A3

Estimates of the Impact of District of Innovation Status on the Percent of Students with Free or Reduced Price Lunch, Disaggregated by Implementation Cohort and Time

Group	Time	Treatment Effect		Lower 95% CI	Upper 95% CI
2017	2013	0.00	*	-0.01	0.00
2017	2014	0.00		0.00	0.01
2017	2015	0.00		-0.01	0.00
2017	2016	0.00		-0.01	0.00
2017	2017	0.00		0.00	0.00
2017	2018	0.00		-0.01	0.00
2017	2019	-0.01		-0.03	0.01
2018	2013	0.00		0.00	0.01
2018	2014	0.00		-0.01	0.00
2018	2015	0.00		0.00	0.00
2018	2016	0.00		0.00	0.00
2018	2017	0.00		0.00	0.00
2018	2018	0.00	*	-0.01	0.00
2018	2019	-0.01		-0.03	0.01
2019	2013	0.00		0.00	0.00
2019	2014	0.00		0.00	0.01
2019	2015	0.00		0.00	0.00
2019	2016	0.00		0.00	0.01
2019	2017	0.00		-0.01	0.00
2019	2018	0.00		-0.01	0.01
2019	2019	0.00		-0.01	0.01

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. 2020 implementers are included in the comparison group for all implementation cohorts. The rows highlighted in gray indicate a treatment effect estimate. The remaining rows are placebo tests of an effect before implementation. Confidence intervals are estimated with bootstrapped standard errors, adjusting for clustering within districts. *p<0.05

Table A4

Estimates of the Impact of District of Innovation Status on Percent of Students with an Individualized Education Plan, Disaggregated by Implementation Cohort and Time

Group	Time	Treatment Effect	Lower 95% CI	Upper 95% CI
2017	2013	0.00	0.00	0.01
2017	2014	0.00	0.00	0.00
2017	2015	0.00	0.00	0.00
2017	2016	0.00	0.00	0.00
2017	2017	0.00	0.00	0.00
2017	2018	0.00	0.00	0.00
2017	2019	0.00	0.00	0.01
2018	2013	0.00	-0.01	0.00
2018	2014	0.00	0.00	0.00
2018	2015	0.00	0.00	0.00
2018	2016	0.00	0.00	0.00
2018	2017	0.00	0.00	0.00
2018	2018	0.00	0.00	0.00
2018	2019	0.00	0.00	0.00
2019	2013	0.00	0.00	0.00
2019	2014	0.00	0.00	0.00
2019	2015	0.00	0.00	0.00
2019	2016	0.00	0.00	0.00
2019	2017	0.00	0.00	0.00
2019	2018	0.00	0.00	0.00
2019	2019	0.00	0.00	0.00

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. 2020 implementers are included in the comparison group for all implementation cohorts. The rows highlighted in gray indicate a treatment effect estimate. The remaining rows are placebo tests of an effect before implementation. Confidence intervals are estimated with bootstrapped standard errors, adjusting for clustering within districts. No treatment effect estimates are significant at 5% significance level.

Appendix B

District of Innovation Characteristics by Exemptions

Table B1

Average Difference in District of Innovation Characteristics by Exemption Status

	Scheduling				Class Size		Certification			Contracts			Student Behavior		
	25.0811	25.081	25.0812	25.082	25.112	25.113	21.003	21.053	21.057	21.102	21.401	21.352	25.092	37.0012	25.036
	First Day	Minutes	Last Day	Pledge	Class Size	Notice Class Size	Cert.	Present Cert.	Notice Cert.	Prob. Contract	Service Days	Teacher Evals	Attend.	Behavior Coor.	Transfers
	N = 850	N = 360	N = 245	N = 215	N = 379	N = 315	N = 756	N = 252	N = 282	N = 453	N = 309	172	N = 228	N = 190	N = 208
Teacher Characteristics															
Teacher Experience	-0.2	-0.03	0.06	0.05	-0.03	0.13	-0.01	0.04	-0.03	0.36*	0.39*	-0.05	-0.72***	-0.38*	0.31
	(0.63)	(0.16)	(0.18)	(0.18)	(0.16)	(0.17)	(0.24)	(0.18)	(0.17)	(0.16)	(0.17)	(0.2)	(0.18)	(0.19)	(0.19)
Teacher Tenure	-0.31	-0.05	-0.16	-0.03	0.04	0.08	-0.27	-0.11	-0.12	-0.1	-0.29*	0.16	-0.31*	0.09	-0.26
	(0.51)	(0.13)	(0.14)	(0.15)	(0.13)	(0.13)	(0.19)	(0.14)	(0.14)	(0.13)	(0.13)	(0.16)	(0.15)	(0.16)	(0.15)
Teacher Turnover	0.01	0.0	-0.0	0.0	0.0	0.0	0.01	0.0	0.01	0.02*	0.02***	-0.01	0.0	-0.01	0.03***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student Teacher Ratio	-0.24	-0.32*	-0.61***	-0.15	0.48***	0.47*	-0.79***	-0.07	-0.29	-0.53***	-1.72***	0.37	0.74***	1.27***	-1.13***
	(0.64)	(0.16)	(0.18)	(0.19)	(0.16)	(0.17)	(0.24)	(0.18)	(0.17)	(0.16)	(0.16)	(0.2)	(0.18)	(0.19)	(0.19)
Student Characteristics															
Percent Hispanic	-0.14*	-0.01	0.01	-0.04*	-0.03	-0.04*	-0.05	-0.0	-0.05*	-0.07***	-0.03	-0.02	0.05*	0.02	-0.04*
	(0.07)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Percent White	0.13	-0.0	-0.0	0.04	0.02	0.04*	0.04	0.0	0.05*	0.08***	0.07***	0.02	-0.08***	-0.05*	0.06***
	(0.07)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Percent Black	0.01	0.0	-0.01	-0.0	0.01	0.01	0.0	0.0	0.0	-0.01	-0.02***	0.0	0.02***	0.02*	-0.01
	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent Econ. Disadv.	-0.12*	-0.02	-0.0	-0.04*	-0.01	-0.03	-0.01	0.0	-0.01	-0.0	0.01	-0.03*	0.01	0.0	0.02
	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
STAAR Performance	0.43*	0.04	-0.01	-0.0	0.02	0.07	-0.05	-0.1*	-0.04	-0.06	-0.1*	0.18***	-0.09	-0.02	-0.06
	(0.19)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)

Note. Table presents the average difference in 2015-16 characteristics between District of Innovation that do and do not exempt each of the most common exemptions. Exemption statistics are as of June 2020. Standard errors are reported in parentheses. *** = p<.001, ** = p<.01, * = p<.05

Appendix C

Average Effects by Grade and Subject

Table C1

Aggregated Estimates of the Impact of District of Innovation Status on School-Level Academic Achievement

Math		**Reading**	
3rd Grade	-0.09	3rd Grade	-0.02
	(0.04)		(0.04)
4th Grade	-0.07	4th Grade	-0.07
	(0.05)		(0.04)
5th Grade	-0.07	5th Grade	-0.01
	(0.06)		(0.04)
6th Grade	0.02	6th Grade	-0.05
	(0.05)		(0.04)
7th Grade	-0.14	7th Grade	-0.07
	(0.06)		(0.05)
8th Grade	-0.01	8th Grade	-0.0
	(0.06)		(0.04)
Algebra	-0.01	English 1	-0.02
	(0.04)		(0.03)

Note. Treatment effects are estimated using a simple difference-in-differences strategy where schools within Districts of Innovation that have not yet implemented their Innovation plan are treated as the comparison group for those that have. 2020 implementers are included in the comparison group for all implementation cohorts. Standard errors are estimating with bootstrapping, adjusting for clustering within districts. After Bonferroni adjustment, no treatment effect estimates are significant at 5% significance level. *p<0.05