Bridging the Gap: Assessing the Effectiveness of Fair Student Funding in New York City Public Schools

Acknowledgements

First, I would like to thank my thesis advisor, Professor Roland Fryer. I've always had an inclination toward public policy, but Professor Fryer was one of the first people to introduce me to the field of data-driven education policy. His class was my first immersion in empirical research, and his intense passion for using data to produce real results and help real people is truly inspiring.

I would also like to extend a tremendous amount of thanks to Professor Greg Bruich for the tireless efforts he put into helping me complete this thesis. Greg has served as a sounding board for initial ideas, a source of inspiration for thesis topics, a treasure trove of knowledge in empirical research and econometrics, and a mentor in the field of economic research writ large.

Many thanks to Ross Lipsky, my high school calculus teacher, dear friend and trusted confidante. Ross has stood by me through this entire process, sometimes taking calls in the middle of the night to hear out my ideas and provide feedback.

A special thanks to Professor James Simpson from the Harvard English Department, whose insight allowed me to arrive at the topic of education funding for this project. I would also like to thank Michael Richard for providing comprehensive comments on this work.

I would like to thank my family for their constant love and support. I especially wish to thank my parents and my brother for their strong belief in public education, fueled by my father's 28 years and counting as a public school teacher. Finally, I wish to thank my friends, all those who kept me sane. I wish to extend a special thanks to Aaron Suduiko and Alex Wang for their encouragement and highly-appreciated camaraderie throughout this process. To all who played any part in this project, from the bottom of my heart, thank you.

I: Introduction

According to UNICEF, a quality education includes content that is related to the acquisition of abilities in literacy, numeracy, and basic life skills. A quality education also includes outcomes that encompass knowledge, skills, and attitudes linked to national goals for education and positive participation in society (UNICEF, p. 3). The right to a quality education is a principle reflected in federal, state, and local statutes across the United States. In New York State, however, this right is not as easily accessible for some families and children as it is for others. On Long Island, for example, the village of Garden City's residents have a median income of approximately $120,000 per year and a public high school that boasts a graduation rate of 97%. Meanwhile, bordering village Hempstead hosts residents with a median income of $45,000 per year and a public high school with severe budget cuts and shortfalls every year, and a much lower graduation rate of 38% (Newsday).

Considering the present educational disparity in New York, the gap was even larger before 2007. The New York State Court of Appeals ruled in *Campaign for Fiscal Equity, Inc. v. State of New York*, after many appeals from 1982 to 2006, that the state's financing system for public education was unconstitutional due to extreme unfairness (Rebell, p. 24). The plaintiff, the Campaign for Fiscal Equality, claimed that NYC public schools received about $400 per pupil less (12% less) than their comparable peers in other parts of the state (Rebell, p. 25). For the state to form an equitable funding structure, the court recommended that the state address the shortcomings of the current system to ensure that every school have the resources necessary to provide a sound basic education. In order to do so, the state needed to create a true foundation formula, or an outline of the basic needs of each school. In addition to a true foundation, the state needed to make sure their budget was sustainable for the future, and needed to develop an accountability system that ensures program quality, equitable distribution of funds, and efficient use of resources (Rebell, p. 25).

In many parts of the state, schools are funded using a mix of local, state, and federal tax dollars, at around a 40-40-20 split. Therefore, 40 percent of school funding comes from local property tax bases, which means that poorer areas cannot fund their schools as well as richer areas due to lower tax bases. This funding split explains much of the difference in budget adequacy between Garden City and Hempstead, and contributed in large part to the inequality that led to the *CFE v. New York* ruling. While the state legislature needed to make changes across the entire state, changes were especially drastic in New York City.

Before the 2007-2008 school year, New York City's Department of Education (DOE) instituted a policy known as Fair Student Funding (FSF), which restructures public school funding to be based on need. FSF allocates money through the New York City Department of Education to schools based on factors thought to be correlated with a necessity for higher funding, such as percentage of students in poverty, percentage of ELL students (English as a Learned Language), percentage of students enrolled in special education programs, and so on. The goal of such a program was to mitigate the inequality that had remained present in New York City's school systems for many years while not immediately taking money away from well-off schools. Ten years after implementation, the policy definitely had a positive effect on the funding levels received by each school in the city, since FSF required funding not to decrease in any school in the city. However, to this point, it is still unclear if the policy shift has produced any measurable changes in student performance, behavior, or educational attainment.

This study uses a quasi-experimental research design to measure the effects of this policy change, comparing changes in performance and behavior in New York City schools with changes at other schools in New York State. This study finds that overall student test scores have only either marginally increased or not increased at all after the implementation of FSF. Graduation rates have generally increased, attendance rates have increased, suspension rates have increased, and future career plans have tended toward more students attending four-year colleges. When results are broken down by subgroup, this study finds that black and Hispanic populations tend to improve by smaller margins than white and Asian populations on ELA exam and State Math Assessment scores. Students with higher socioeconomic status (SES) improve by larger margins than do students from lower SES backgrounds. Disabled students improve their test scores by far larger margins than their general education counterparts, and student not proficient in English improve their 4th grade reading and math scores more than their English-proficient peers. However, in a strangely conflicting result, graduation rates for the disabled decrease substantially as well. All regressions are estimated with and without allowing school-specific differential linear time trends to account for preexisting differential trends across schools. However, the overall conclusions are still not definitive. These results are robust to changes in research design: propensity score matching and DiNardo, Fortin, and Lemieux (DFL) reweighting produce very similar results in every outcome variable. The improvements in test scores are promising, but the results of this analysis provide little solid foundation on which to conclude that FSF is effective in its mission to improve performance of less fortunate students.

The remainder of this paper is structured as follows: Section II reviews the previous literature regarding the link between education and poverty, as well as the current problem of educational inequality and previous attempts to fix the problem in other jurisdictions. Section III examines the reason behind the change to FSF. Section IV outlines the basic framework for the Fair Student Funding program. Section V reviews more previous literature examining whether

FSF actually succeeds at redistributing funding across NYC public schools. Section VI discusses the data used in this study. Section VII outlines the ideal experimental design for this study in order to establish the ultimate goal of this research. Section VIII describes the reasons that the ideal experimental design is infeasible. Section IX discusses the experimental design used in this study. Section X enumerates the conditions under which the results of this study can be interpreted as causal effects. Section XI analyzes the validity of this study's assumption of parallel trends. Section XII provides overall results for test scores, graduation rates, attendance rates, suspension rates, and post-high school aspirations. Section XIII breaks down test score results by race. Section XIV breaks down test score results and graduation rates by subgroups other than race, including SES, English proficiency, gender, and special education status. Section XV unpacks these results and summarizes the most important takeaways. Section XVI uses two different research designs, propensity score matching and DFL reweighting, to perform robustness checks. Section XVII outlines some limitations of the data and the research design of this study. Section XVIII discusses some policy implications of these findings. Section XIX suggests future research paths in this field. Section XX concludes.

II: Literature Review

What determines educational outcomes?

Education and poverty are generally inversely related across the world, in both developing and developed countries. Tilak (2002) uses a meta-analysis to review the role of education in alleviating poverty in developing nations, and finds that education and poverty form a positive feedback loop between themselves. If a lack of education funding presents itself, a vicious cycle can emerge, in which lack of education leads to more poverty, which then leads to less funding for education, which then leads to more poverty (Tilak, p. 198). This cycle is based on an empiricallybacked assertion: the more educated a population, the smaller proportion of that population in poverty. Likewise, if education funding is sufficient, there will be poverty reduction, and so on.

Tilak (2002) also focuses on common misconceptions regarding education funding in poor areas; some include the notion that poor societies cannot afford good education, primary education is enough to escape poverty, the state is not responsible for educating its citizens, and a qualified teacher is not important in basic education. Tilak (2002) finds based on previous literature that these approaches, which cover adult education, decentralization, and privatization, might actually be harmful to the development of a solid education system (Tilak, p. 202).

The relationship between education, poverty, and housing is also important when considering public policy. Schwartz et al. (2010) compares the education received by children in public housing versus that of children in regular housing. Public housing is located in urban areas in which crime rates are high, median income is very low, and the community is majority-minority. The study begins by figuring out whether or not students who live in public housing perform worse on standardized tests than do students who do not live in public housing, which it empirically finds that they do. Studying academic performance from the 2002-2003 school year, the average fifth grade student living in public housing scores 0.31 standard deviations below the citywide mean on math tests and 0.33 standard deviations below the city average on reading tests. Meanwhile, students not in public housing score about 0.06 standard deviations above the citywide mean on both math and reading tests (Schwartz et al., p. 76).

However, the real message of Schwartz (2010) comes with the fact that differences in school do not fully explain the difference in test scores between students in public housing versus those not in public housing. The study finds that students living in public housing earn lower scores on standardized tests on average than their schoolmates who attend the same school but live outside of public housing. Specifically, among non-poor students (students not eligible for free or reduced-price lunch), the difference in reading and math between public housing and non-public housing students' test scores hovers just under 0.3 standard deviations, and among poor students, the difference is 0.15 standard deviations (Schwartz et al., p. 82). This finding suggests a notion that has been written about in other studies, but Schwartz backs it up with data: there is another unobservable factor that contributes to educational success, and a further analysis of community and home environments will be needed to find that effect. In addition, there is a need for more nuanced analysis of school policies that might enable schools to serve students in public housing better than they currently do (Schwartz et al., p. 83).

Chetty, Hendren, and Katz (2014) examines a similar idea by using data from the Moving to Opportunity (MTO) experiment, which gave randomly-selected low-income families housing vouchers to move into lower-poverty neighborhoods. The idea of the MTO experiment is to assess the impact of housing on children's long-term outcomes. The MTO experiment study finds that moving to a lower-poverty neighborhood increases college attendance rates, increases earnings rates, and decreases single parenthood rates, whereas moving as an adolescent results in slightly negative effects due to disruption. Overall, MTO supports the notion that housing and neighborhood characteristics may have a significant effect on the educational and wage-earning outcomes of children (Chetty, Hendren, and Katz, p. 856). The reduced-form intent to treat (ITT) effect of moving into a low-poverty neighborhood is a 14% increase in average future wages and a 16% increase in college attendance over the comparison group (Chetty, Hendren, and Katz, p. 857).

Experts can disagree about the correlation or the causation between socioeconomic status (SES) and educational outcomes, but overall evidence suggests that students who come from lower

SES backgrounds experience worse educational outcomes. Ladd (2012) analyzes the average difference in standardized test scores between the top ten percent and the bottom ten percent on the income distribution. Ladd finds that the difference between the top and bottom deciles is 0.7 standard deviations for those born in 1945. When we consider the birth cohort from 2001, that split increases to 1.2 standard deviations. The income gap during this time has almost doubled while, over the same time period, the black-white gap has decreased from 1.25 to 0.7 standard deviations (Ladd, p. 205).

Much of the literature focuses on early childhood as a crucial time for improving educational outcomes. Through a meta-analysis of 38 studies, Barnett (1998) finds that early childhood education provides increases in cognitive development for poor children, but the longevity of these effects is less certain. Barnett finds that early childhood education produces persistent positive results in achievement and academic success, but does not provide the same persistence in IQ. Head Start and early development programs in public schools provide these same effects as well. Cost-benefit analysis of these programs shows that the benefits of Head Start far exceed the costs because of the increase in productivity and social responsibility that will eventually become apparent later in life. Barnett maintains a policy recommendation of at least one year of quality education prior to school entry, whether that means part-day preschool education or a full-day developmental program (Barnett, p. 205).

Many studies also hold an important role for parental education in the developmental process. According to Cooper et al. (2009), past research has indicated that poor children start school with far lower cognitive skills than rich children, and the gap widens as school goes on. The purpose of Cooper (2009) is to see if poverty-related differences in parental involvement in education contributed to these differences and whether the academic role of parental involvement varied across race. Home-learning activities didn't explain the lower achievement rates of poor students, but home-learning activities related to math and science seemed to mediate the association between family poverty and kindergarten achievement (0.2 standard deviations) (Cooper et al., p. 870). Overall, the inequality in education heavily depends on income from the start of life. However, the educational gap at the start of school is not just about finance: parental involvement does play a role, but the size of that role likely depends on race. Positive parental involvement appears to be highly statistically significant across the board, but poverty and lack of parental involvement appear to be more significantly negative for black, Hispanic, and white families ($p<0.01$) than they are for Asian families ($p>0.1$) (Cooper et al, p. 874).

Fryer and Levitt (2015) tests the effectiveness of parental involvement by providing financial incentives for parents to engage in behaviors believed to increase early cognitive development such as attending early childhood sessions and completing homework assignments with their children. The effects of this program on cognitive skills appears to be large and positive but statistically insignificant (0.12 standard deviations with standard error 0.094), while non-cognitive skills see a large, positive, and statistically significant effect (0.203 standard deviations with standard error 0.083). The study then examines heterogeneity across racial lines, and finds that (consistent with the literature), white and Hispanic students see large statistically significant increases in cognitive and non-cognitive skills while Black students see a negative but statistically insignificant effect. Whites see cognitive skills increase by 0.932 SD (0.353) and non-cognitive skills increase by 0.821 SD (0.181), while Hispanic students see cognitive skills increase by 0.367 SD (0.133) and non-cognitive skills increase by 0.428 SD (0.122). Fryer and Levitt (2015) agrees with the already-present literature regarding the importance of parental involvement in child development.

Other studies find the relationship between parenting quality and poverty to be tenuous at best. Dermott and Pomati (2015) holds a big role for parenting in a child's upbringing, but does not find that the difference between rich and poor families is significant. Dermott and Pomati (2015) is based on UK data, but is generalizable to the rest of western culture due to its focus on what is popularly considered to be 'good parenting.' Western nations such as the US and the UK consider good parenting to be a sort of "concerted cultivation," as opposed to the "natural growth" philosophy pushed in previous generations (Dermott and Pomati, p. 126). Dermott and Pomati (2015) finds that practices such as reading, playing games and eating meals together do not happen less frequently among poorer families, and therefore are not part of the difference between rich and poor children. Despite the fact that parenting has been pushed as a main location in which we can invest time and money to boost social mobility, claims that poor parents are poor at parenting are not driven by empirical evidence (Dermott and Pomati, p. 134).

It is also important to consider the role of reverse causality in this relationship between education and poverty. Buck and Deutsch (2014) postulate that it may be natural to think that poor education causes more poverty, but it is also possible that more poverty causes worse educational outcomes. The main focus of Buck and Deutsch (2014) is on the effects of a poor community on the local public school. External effects of a bad neighborhood can contribute to a child's lower level of school achievement through limited parental employment, stress, poor nutrition, crime, and so on (Buck and Deutsch, p. 1141). In turn, the inability to escape poverty can lead to a defeated attitude among an entire community, leading students to believe they are "being left behind due to a macro-level failure, which influences micro-level behavior" (Buck and Deutsch, p. 1143). When this macro-level breakdown of the school system breaks individual students they become far less able to cope with life's future problems. Behavioral changes can lead to increased disciplinary issues, leading to detentions and suspensions. In this event, students and teachers cannot build meaningful trustworthy connections, which makes quality education even harder to accomplish.

New York State funding structures were previously inadequate

In response to the *CFE v. State of New York* ruling, the state legislature made wide-reaching efforts to ensure that all schools received what the New York State Court of Appeals called "the opportunity for a sound basic education" (Rebell, p. 24). According to Baker (2014), New York State had one of the least equitable funding structures in the country for decades leading up to 2007, with poorer areas across all types of communities (urban, suburban, rural, etc) receiving far less money than was deemed acceptable by the state. The year 2007 brought on myriad plans for funding changes, including competition among local school districts for additional funds above 'basic needs,' or what the state determined to be a minimum funding level necessary to ensure a sound quality education. The state made what Baker believed to be good targets for what funding should look like, and then did not adequately fund the programs. Basic funding levels were known as "foundation aid" levels (Baker 2014, p. 7). When implementing the foundation aid formula, the state considered state aid per pupil to be one of its most important metrics, and Baker found that the highest-need districts fell short by almost $4,000 per student.

New York Governor Andrew Cuomo instituted property tax caps soon after *CFE v. New York State* to decrease overspending and inefficiency, but it has been found that tax and spending limits usually result in service quality reduction rather than efficiency gains, inequality generally rises when these provisions are given voter overrides (as were given in this case), and tax relief often leads to increased inefficiency. New York State has engaged in systemic underfunding of its own plan to fix inadequate funding. It needs to fully fund its plan if it wants to fix the problem (Baker 2014, p. 13).

Other jurisdictions have had to deal with unequal school funding before

Previous attempts have been made across different jurisdictions to deal with unequal funding. Escue (2012) analyzes the funding structure of Florida public schools, which is currently similar to the funding structure of New York City public schools *before* FSF. Currently, Florida employs the concept of horizontal equity in their school funding structure, in that equal students receive equal funding. However, the state is trying to implement vertical equity as well, the notion that unequal students receive proportionately unequal funding. This style of funding is difficult to implement due to its reliance on value judgments and priorities when determining how students are 'unequal' (Escue, p. 348).

Currently, the state of Florida has only 67 school districts serving a population of over 19 million people. For reference, Long Island alone has 127 school districts serving just under four million people, resulting in an average district size of just under 32,000 people. Larger districts often signify less-targeted policy: Miami-Dade school district, covering the entire city of Miami and some surrounding suburbs, serves the wealthiest school and the most impoverished school in the state; yet, the current Florida funding system equalizes state aid across the entire district. Escue attempts to rectify this problem by introducing a poverty index model to try to increase the effectiveness of Florida's school funding. The plan currently uses incidence of free and reduced lunch per school along with other financial indicators as proxies for financial need, but Escue aims to find a better model (Escue, p. 349).

Escue does come up short, however, in that it suggests and builds a school-level model but does not perform any analysis on Florida's education policies. The model should be expected to perform at least as well as the current policy regime, but the study would benefit at least from a simulation of some kind. Despite all that, a poverty index model would be a good place to start for any state attempting to overhaul its education funding model (Escue, p. 367).

The federal government has tried to address the problem of teaching kids from different SES backgrounds as well. Ladd (2012) considers the position of No Child Left Behind (NCLB) in our national conversation on education. According to Ladd, one large problem with No Child Left Behind is that it ignored the difference in starting points between children from different SES backgrounds, and required that all children reach the same benchmarks for each grade. This has led to the failure of a large number of schools, narrowing of the curriculum, and low morale among teachers who believe they cannot achieve the goals set for them (Ladd, p. 204).

Baker (2009) looks at a wide range of policies that broadly fall under the concept of 'weighted student funding' (WSF), of which FSF is an example, which puts higher weights on school characteristics that would indicate a need for more funding. For example, a high weight is placed on ELL, or English as a Learned Language. Therefore, if a school has a high percentage of students for whom English is not their native language, that school will receive more funding for English-learning programs. This article makes clear that the goal of WSF, and by extension FSF, is to target resources to higher-need schools, but also that much of the eventual resulting funding distribution is politically driven. While describing an outcome, Baker states "such an option is typically a political concession to 'spread the wealth' rather than target resources to higher-need schools—the espoused goal of weighted student funding" (Baker 2009, p. 7). Baker (2009) finds, using a two-stage least-squares regression, that schools in Texas and Ohio that use WSF do not receive more equitable funding, and therefore cannot be distinguished from other school districts that fund their schools in the more traditional method.

Even when some jurisdictions tried to address the issue of unequal school funding, unfavorable events transpired to make reform even more difficult. The 2008-2009 recession greatly hindered New York State's ability to fund its schools appropriately. Since most schools in New York are funded through local property taxes, the housing market crash that precipitated this financial crisis caused a decrease in property tax revenue of 13.5% over the course of the recession (Chakrabarti and Setren, p. 4). Local and state governments faced steep budget cuts due to this revenue shortfall, so many of them had to cut non-core subject programs such as extracurriculars, art and music, maintenance, and transportation, among others. The American Recovery and Reinvestment Act (ARRA), commonly known as the 2009 economic stimulus package, provided about $5.6 billion to New York schools to help stave off some of these budget cuts. However, the result still involved heavy cuts in state aid, especially in some of the highest-need school districts in the state such as Buffalo, Rochester, and Syracuse (Chakrabarti and Setren, p. 6). The main effect the recession had on New York state public schools is not so much that total school expenditure changed, but rather that the composition of funding changed. Reliance on federal funding increased dramatically, while shares of state and local funding dropped when ARRA funding began. Since ARRA funding went to higher-need districts first, the result was a more equitable redistribution of state education funds. However, many school districts around the state had to slim their budgets, making this change harder to handle for many school districts (Chakrabarti and Setren, p. 19).

Before Fair Student Funding

New York City public schools used to be organized into 32 different "community school districts," each overseen by a superintendent, not unlike many school districts in the rest of New York State. Superintendents would be given funds based on the number of students in each district, and would then allocate money to each school within their respective districts. Before the 2002-2003 school year, Mayor Michael Bloomberg was given control of New York's public schools, and budgeting became increasingly centralized under the Department of Education, which allocated funds directly to schools rather than districts (IBO 1, p. 2). Even after the New York City Department of Education took central control of budgeting, there was still a rather unequal distribution of funds between schools in different parts of the city. Analysis conducted by the New York City Independent Budget Office (IBO) in 2005 uncovered large disparities in per student spending for general education classrooms. Based on this system, in theory, every school should spend approximately the same amount per student. However, the IBO study found that classroom spending varied from a low of $2,511 to a high of $8,569 with funding levels spread out relatively evenly across this range (IBO 1, p. 1).

Within this overall result, the study finds that elementary and middle schools with higher student spending generally have fewer students per teacher, higher teacher salaries, smaller class sizes, and lower enrollment. However, there was little evidence to suggest that indicators of student needs, such as the share of English Language Learners (students for whom English is not a native language), were significant factors in predicting funding per student, even after controlling for teacher staffing and school size. Based on these findings, the city leadership felt it needed a new formula to better allocate funds to schools that truly need them.

III: Why Make a Change?

The idea for a change in policy stems from two main issues with the previous system. First, the old system generally aimed to provide equal funding for all schools; however, certain schools have more high-need students than others, and those schools would need more funding to provide a similar level of educational quality. Second, even if we were to accept the goal of equal funding for all students, the previous system did not even achieve that objective; the difference between the schools with the lowest to highest per pupil funding was over $6,000, often with higher funds skewed toward lower-need schools (IBO 2, p. 2). According to the DOE, two schools with similar student populations (which should have received approximately the same amount of funding) could show a difference of $1.5 million in funding per year in the early 2000s, approximately half the yearly budget of a well-off school (NYC DOE, p. 2). Therefore, the goal of equal spending for all students was not even met, and the program proved to be somewhat regressive in practice.

IV: Fair Student Funding

The goals of a policy change in this realm were threefold: increase equity while preserving stability, improve student achievement, and make school budgets more transparent (NYC DOE, p. 2). To achieve these goals, starting in the 2007-2008 school year, the DOE switched its funding structure to a formula that uses a weighted, needs-based approach to making school budget decisions called 'Fair Student Funding' (FSF) (IBO 2, p. 2). First, according to the DOE, FSF increased equity in funding while preserving stability by directing new funds toward needy schools without taking away funds from other schools. The new infusion of money came from the DOE's reduction of central and regional offices (NYC DOE, p. 2). Second, FSF increased student achievement by creating financial incentives for schools to take in more struggling students by awarding more funds to those schools, and rewarding schools that improve the results of those students. Finally, FSF aimed to simplify funding streams and make school budgets more transparent; before FSF, budget choices were often hidden in the fine print, but this new system created a weighted formula that is public knowledge, so all school funding allocation could be followed and understood by the public (NYC DOE, p. 2).

First, every school started with a base amount of $200,000 to be used at the principal's discretion. Past that, the system uses weights for factors thought to be positively related to higher-need institutions (IBO 2, p. 2). Every school receives a grade weight, which is applied to the number of students in elementary school (K-5), middle school (6-8), and high school (9-12). High school students are weighted higher than elementary school students because high schools offer electives and require more non-personnel costs (usually more expensive science equipment). Middle school students are weighted highest because middle school often sees the largest academic drop-off out of any time period in a child's schooling (NYC DOE, p. 4). Students who need more academic intervention (children who come into the school with a high chance of failure) are weighted higher, along with English Language Learners (ELL) and Special Education students. Career and technical schools, specialized academic schools, specialized audition schools, and transfer schools also receive more money (NYC DOE, p. 4). FSF did not take any money away from schools during its first few years in order to preserve stability, allowing needy schools to increase funding without taking funds away from well-off schools.

The funding floor, guaranteeing that no school would be worse off as a result of FSF, came as a result of political compromise. Joel Klein, chancellor of New York City public schools at the time FSF was implemented, knew it would be difficult to put such a reform in place. The United Federation of Teachers (UFT) protested the stipulation in FSF of tying teacher salary to the budgeting process. Parents were worried that good teachers would leave their children's schools because of FSF (which was often true in order to redistribute good teachers into more schools that needed better teachers). This alliance between the teachers' union and well-to-do parents proved to be highly difficult to handle, so the eventual zero-sum game between good schools and bad schools would be phased in, rather than being implemented all at once. In Klein's words, "equity can be costly" (Klein, p. 277).

V: Does Fair Student Funding Work?

One of the DOE's stated goals in implementing FSF is to improve student performance. However, there has been little empirical research regarding whether or not academic achievement rates of students in New York City have actually improved. This study aims to use data from New York State Education Department to compare student achievement between NYC and non-NYC public schools before and after the 2007-2008 school year. Student achievement will be measured in terms of English Language Arts (ELA) and State Math Assessment test scores for 4th and 8th graders, high school graduation rates, and post high-school plans (future aspirations, going to college, going straight into the workforce, enlisting in the military, etc.). In addition to student performance, this study will also measure changes in student disciplinary records in response to this change in funding, measured through attendance and suspension records by school. After overall effects are examined, this study will then split the data into subgroups by race, gender, and socioeconomic status (SES), among other factors, and determine if outcome variables have significantly increased in any of these subgroups. This study aims to find some conclusive evidence that can speak to the effectiveness of FSF in New York City and its potential for use in other cities across the country.

VI: Data

The New York State Education Department (NYSED) has compiled school-level and district-level data extending from the 1999-2000 school year to the 2015-2016 school year on ELA and State Math, test scores, high school graduation rates, attendance records, suspensions, and aspirational goals (college, military, work after high school), among others. The data are split by race, gender, SES, immigrant status (immigrant vs. not immigrant), English proficiency, and presence of disabilities. This study only uses data starting in the 2001-2002 school year due to a lack of subgroup data in the first two years of the sample. These data exist for all public schools in New York State, and each school has its own unique ID number, enabling this study to differentiate between NYC and non-NYC schools (and therefore between treatment and comparison units) with relative ease. The ELA and State Math exams are state standardized tests, so all schools in this study will be compared on the same academic scale.

It appears that NYSED did not normalize its test score data format until the 2006-2007 school year, because each year from 2001-2002 to 2005-2006 has a different method through which the data was formatted. In earlier years, subgroups were in their own columns rather than their own rows, and almost all variables had different names. The first step of this project was to operationalize these data by constructing a panel data set with 4,650 schools, 13 years, and up to 19 different subgroups for a total over 1.1 million observations.

Test scores serve as a proxy for school quality because they do not follow the same students every year, which means the quality of any individual student cannot affect these test scores over time. This leads to an assumption that persistently high test scores signify a high-quality school, and change in test scores lead to insights on changes in school quality. The test results are split into four levels of increasing academic performance, aptly named Level 1, Level 2, Level 3, and Level 4. Levels 1 and 2 are failing grades, and levels 3 and 4 are passing grades. Since increases in levels 2 and 3 can be attributable to decreases in the levels above them *or* below them, their coefficients can lead to ambiguous information. This study omits these results due to this ambiguity. Therefore, this study measures the effect of treatment on a school's percentage of students scoring at level 1, the lowest level (unambiguously bad), the percentage of students scoring at level 4, the highest level (unambiguously good), and the percentage of students scoring at levels 3 or 4 (the percentage of students passing the test).

Graduation rates measure the percentage of seniors in each high school who earn a high school or other equivalent diploma, attendance rates measure the average percentage of enrolled students who attend school on a given day, and suspension rate measures the number of suspensions per student in a given year. Finally, aspirations measure the percentage of high school graduates who plan to enroll in two-year or four-year colleges, enlist in the military, enter the workforce, and so on.

VII: Ideal Experimental Design

If it were feasible, the ideal method this study would use to examine the effect of FSF on student performance is to run a randomized experiment. We would randomly select 100 schools in New York City, randomly assign 50 of those schools to change their funding structure to FSF, and leave the other 50 schools with enrollment-based funding. This experiment would resample treatment and control groups until it achieved covariate balance, as per the Imbens and Rubin (2015) model. Imbens and Rubin (2015) discusses matching on covariates and rejecting matchings with poor covariate comparisons. Then, over a period of between 10 and 15 years, long enough to allow the effects of the funding change to manifest themselves, the study would measure test scores, graduation rates, college acceptance rates, dropout rates, attendance rates, and so on.

The experimental design would regress post-treatment outcomes on a treatment group indicator variable. Since we are ensuring covariate balance, a difference in difference regression is unnecessary, and we can instead use a simple OLS regression. The covariates we would consider in this ideal experiment would be demographic factors such as race and gender, average family income, outcome variables from years *before* the treatment year such as pre-treatment test scores and graduation rates, and crime rates, among others. This treatment would not allow students to leave the public school system or change schools at all during their tenure in order to track similar students through test scores in various grades and through graduation rates. This is an ideal research design because treatment assignment would be unrelated to any other determinants of student performance, including observables such as the aforementioned covariates and unobservables such as teacher quality, among others.

There may be concern that using outcome variables from pretreatment years would cause bias in regression coefficients. In most cases, any variable that occurs before treatment is considered a covariate, not an outcome, but variables in panel data often exhibit autocorrelation over time. Therefore, variables used as outcomes after treatment might be associated with values before treatment. However, this is not a concern in this idealized experiment due to the randomized nature of the treatment assignment. The fact that covariate balance includes pretreatment outcome variables would not make use of improper control variables because this experiment would deal with panel data, measuring the same schools over time. Using a randomized treatment assignment scheme would allow for pre-treatment test score balance among the treatment and control populations.

VIII: Ideal Experiment is Infeasible

This type of randomized experiment is not possible because this type of public policy changes should not be conducted by random lottery. If a governmental entity wishes to make a policy change, it often finds who is eligible and permits everyone in that population to access the reform. Depriving students of a potential new method to help them succeed is often not assigned by lottery due to both the perception and potential existence of unfairness in location. Some papers, such as Baicker et al. (2013), do use this random lottery when considering expanding policies such as Medicaid, but this does not extend effectively to a policy intended to affect an entire region's public schools. Baicker et al. (2013) uses Oregon as a natural case study to see how a new form of Medicaid expansion affects the health, healthcare utilization, and financial outcomes of those who use it, but those who are not selected in the lottery still retain the old form of Medicaid, widely thought of as a sufficient form of health coverage for poor Americans. Chetty, Hendren, and Katz (2015) use a similar research design using data from the Moving to Opportunity project, which randomly selected families to move into lower-poverty neighborhoods. Similar lottery designs can be found in Fryer (2013), Jacob and Ludwig (2012) and Bettinger and Slonim (2006) as well.

However, in the case of this study, those whose schools are not selected to partake in FSF would be relegated to the status quo, which has been deemed by the New York State Court of Appeals as unsatisfactory to ensure every child a quality education. Therefore, a lottery is not feasible in this circumstance.

Another reason this ideal experimental design is not feasible is that there is no way to hold all else constant while FSF is allowed to run its course. So many different conditions could change over the course of the study that would affect a child's education and a school's performance that we could never perform such an analysis in a vacuum. Economic conditions, testing regulations, and school populations would need to remain constant to isolate the causal effect attributed to FSF with absolute certainty. Some of these confounding factors such as testing regulations, would be differenced out in the experiment due to their presence in the control group and the treatment group. However, this ideal experiment is not feasible in the real world, where it is not possible to control for all these potential confounding variables. In Section X, this study will outline the conditions under which the current research design is able to isolate the causal effects of FSF on student performance.

IX: Research Design

This study uses parallel trends as its identification assumption, so this study works under the assumption that both NYC and non-NYC schools follow the same trends before the treatment year. This assumption brings us closer to being able to estimate causal effects from the resulting regression coefficients. Section X will outline conditions under which these regression coefficients can be interpreted as causal effects. The treatment group consists of all groups served by New York City public schools, and the comparison group is all groups served by all other public schools in New York State. The difference in trend line over the next seven years can be interpreted as the causal effect of FSF on student achievement under conditions outlined in Section X. After the overall effect has been established, this study will separate the overall data into subgroups by race, gender, SES, and other factors, and run the same difference in difference test on those data subgroups. Standard errors are clustered at the school level for all regressions. Bertrand, Duflo, and Mullainathan (2004) states that we should cluster standard errors at the level of the policy change to avoid inconsistency, but the NYC DOE blurs the line between schools and school districts, so this study also tries clustering the data at the school district level. Results do not change in any significant way.

This study uses three different regressions: a basic difference in differences regression, a regression with year fixed effects and school fixed effects, and finally a regression with year fixed effects, school fixed effects, and school-specific linear time trends. The equations are:

$$y_{st} = \alpha + \beta_1 DiD_{st} + \beta_2 NYC + \beta_3 post + \varepsilon_{st} \tag{1}$$

$$y_{st} = \alpha + \beta_1 DiD_{st} + \gamma_t + \delta_s + \varepsilon_{st} \tag{2}$$

$$y_{st} = \alpha + \beta_1 DiD_{st} + \gamma_t + \delta_s + \lambda_{st} + \varepsilon_{st} \quad (3)$$

In Equation 1, DiD_{st}, which is equal to *NYC*post*, is the difference in differences indicator, *NYC* is the indicator for a New York City public school, and *post* is the indicator for a data point after the 2007-2008 school year. In Equation (2), γ_t represents time fixed effects, and δ_s represents school fixed effects. Finally, in Equation (3), λ_{st} represents school-specific linear time trends. The non-DiD_{st} indicators in Equation 1 and the fixed effects in Equation 2 are collinear. For this reason, *NYC* and *post* have been dropped from Equations 2 and 3. The third regression includes school-specific linear time trends, which allow us to relax the parallel trends assumption by allowing each school to have its own trend. When considering later results, we will only consider regressions 2 and 3 due to the relative lack of information for which regression 1 controls.

The purpose of including school and year fixed effects in this study is that the amount of potential confounding factors is too numerous to include in any tractable regression. Therefore, without fixed effects, this study would be vulnerable to omitted variable bias. For example, this study does not address the quality of neighborhood in which students grow up. The study does not address parental involvement or competence. The study does not address teacher quality. These are just a few variables that are left out of the study due to lack of available data. Year and school fixed effects can help mitigate the influence of omitting these variables from consideration because the aforementioned factors that vary across schools do not change very much over time, and the factors that vary over time do not change much across schools.

The purpose of this process is to determine if FSF has any significant positive effect on the entire population or on any of its subgroups. If FSF does have a significant effect on the overall population, we can then suggest that the policy is effective, and we might consider expanding the policy to more locations. If there is no significant effect, we cannot conclude anything solid, but we will most likely be able to suggest one of three possible conclusions: either FSF is actually ineffective, FSF in its current implementation is not large enough to have an effect, or FSF has not been in place for long enough to see any effects yet. From these conclusions, the policy response to a statistically insignificant effect may not be to remove the program.

X: Causal Effect Conditions

Given that a randomized experiment with perfectly observable causal effects is not feasible in current circumstances, this section addresses the reason this research design gives us the best chance to isolate causal effects. Specifically, this section focuses on conditions under which this study can produce causal effects. The most important assumption this research design must make is parallel trends; in support of this assumption, NYC schools and non-NYC schools exhibited the same (or at least similar) trends in relevant outcome variables before the treatment year. This assumption maintains the notion that, without FSF, NYC and non-NYC schools would have improved at the same rate. This assumption is accounted for in regression equation (3), with the inclusion of an interaction term between year and school fixed effects. Figures 1-12 suggest that this is a valid assumption to make in this model; 4th Grade Math scores are generally not parallel, but all other test scores exhibit parallel trends before the treatment date. Therefore, it is reasonable to assume that the data in this study exhibit parallel trends before the treatment date.

A beneficial feature of a difference in difference regression with fixed effects and linear time trends is that it does not require covariate balance. Propensity scores, which this study uses as a robustness check, would need some degree of covariate balance between treatment and comparison groups in order to find matched pairs similar enough to perform propensity score analysis, but that is not the primary research design of this study. The reason a difference in differences regression is the best choice for this study is that it only requires an assumption of parallel trends, and does not require covariate balance between treatment and comparison groups.

XI: Summary Statistics, Covariate Balance, Linear Trends

Before this study analyzes overall regression results, we will check the strength of covariate balance between NYC school and non-NYC schools before the treatment date, as well as the validity of our assumption of parallel trends. The importance of looking at covariate balance is that it makes the parallel trends assumption more believable. First, the study will look at covariate balance of demographic factors in the year 2007, the year before the treatment took place. Simply put, these covariates are not balanced. NYC schools have far higher minority representation; for example, NYC schools an average of 40% Hispanic students in 2007 compared to just under 9% in non-NYC schools. Only 24% of non-NYC public school students are on a free lunch program, compared to 76% of NYC students. Around 20% of NYC students are not English-proficient, compared to just 3.5% of non-NYC students. The only two 2007 demographics that track closely together are population stability and school size. The percentage of students in 2007 who were in the same school at any point in the previous school year averaged 85% for NYC schools and 93% for non-NYC schools, and the average NYC school held 669 students compared to 570 in non-NYC schools. These differences do not invalidate the research design as long as the rate of change in these covariates over the years in this study does not change.

Next, this study will examine the balance between outcome variables of NYC and non-NYC schools before treatment (Table 2). Since this is a panel data set, pre-treatment instances of future outcome variables are not considered outcomes and can therefore be analyzed as covariates. Across the board, Level 1 test scores seem to be somewhat similar between NYC and non-NYC schools (8th grade Math differs by about 15 percentage points, but that is the largest difference). Level 4 scores seem to be even more similar across the board, and passing scores seem to be fairly far apart (most differences are more than 15 percentage points). Pre-treatment attendance and suspension rates are fairly similar, but graduation rates are widely apart (38.9% for NYC schools and 79.4% for non-NYC schools). These differences in covariates do affect previous time trends in some outcome variables; however, the trends in the outcome variables are still linear, which therefore allows the school-specific linear time trends included in Equation 3 to account for the difference in covariate balance.

For almost any other research design, the lack of covariate balance between NYC and non-NYC schools would constitute a major flaw in the conclusions that can be drawn from the data. However, this lack of covariate balance is exactly why this assumption of parallel trends is so important. A difference in differences regression does not need covariate balance as long as changes would have been the same in the absence of the treatment. We will look at Figures 1-12 before the treatment data to test this assumption of parallel trends and see how closely average test score values track together in NYC and non-NYC schools. The common trend in this set of graphs is that NYC and non-NYC schools move in the same general direction, and can be seen as parallel in most cases. Only a few cases are not parallel, such as 4th grade Math scores and 8th grade Level 1 Math scores. In these cases, NYC schools are closing the gap even before the treatment date. However, in all other test score data, the trends are approximately parallel. While this comparison is not quantified, the evidence suggests that parallel pre-treatment trends do hold in many of our outcome variables, and can likely result in regression coefficients that can be viewed as causal estimates.

XII: Overall Results

XII(a): Nonparametric Evidence

This study begins with simple nonparametric evidence before it delves into regression results. Figures 1 through 12 show previous mean trends in all three test score outcome variables (Level 1 scores, Level 4 scores, and passing scores) for all four test scores in question (4th grade ELA, 8th grade ELA, 4th grade Math, and 8th grade Math). Overall, preliminary examination on mean scores alone suggest that FSF has done a fair amount to improve NYC test scores and graduation rates relative to non-NYC measures. In many cases, both NYC and non-NYC schools track closely together. However, in some cases such as 8th grade ELA passing scores, NYC schools seem to have closed the gap between the two groups. These test scores may have changed their standards sometime around 2012, which slightly changes the metrics and might account for the sudden jumps in the data. However, this does not affect the regression coefficients because the changes affect the entire state, not just NYC schools.

The results discussed below will consider two regression models: Equations 2 and 3 in the experimental design section, corresponding to the fixed effects model and linear time trends model respectively. The reason this study considers those models is that the introduction of linear time trends enables the study to determine whether a certain effect can be explained by preexisting within-school differential time trends. Put simply, the difference in coefficients between these models enables us to see if an effect is due to FSF, or if the schools may have improved anyway even if FSF were not implemented. The coefficients reported in the "Tables and Figures" section of this paper are reported as percentage point changes in outcome variables. However, to explore the relative effects of FSF on student performance, this study will additionally report findings in terms of standard deviation changes relative to the pre-2007 mean.

Of the 364 coefficients in the tables in Section XXII, 243 are significant at the 1% level. Given the high number of outcome variables being tested, it is possible that these coefficients are significant by chance. Bonferroni corrections are conducted to adjust for multiple hypothesis testing. In particular, the standard critical values for 5% significance level are replaced with critical values that correspond to 0.05 divided by the number of regressions (364), so that a p-value must be below 0.00014 to be significant at the 5% level. This increases the standard for significance while removing speculation that these coefficients are significant by chance. After the corrections, 211 of the 243 coefficients originally significant at the 1% level retain their significance at the 0.014% level.

XII(b): ELA Test Scores

The first type of outcome variable we will consider is student test scores, specifically on the 4th grade and 8th grade ELA and State Math Assessments. On the 4th Grade ELA Exam, using time and school fixed effects, there is a 3.05 percentage point (0.32 SD) decrease in Level 1 scores, a 6.18 percentage point (0.48 SD) increase in Level 4 scores, and a 9.13 percentage point (0.46 SD) increase in passing scores. All these findings are statistically significant at the 1% level, and the full tables can be found in Table 3. Using school-specific linear time trends, the effects change by a substantial margin. There is now a 0.23 percentage point (0.02 SD) *increase* in Level 1 scores, but this value is statistically insignificant at any reasonable level, and the 95% SD confidence interval (CI) rules out any decrease larger than 0.05 SD. There is still a positive and significant effect on Level 4 scores, but the point estimate drops to 0.64 percentage points (0.05 SD) and the 95% CI rules out any increase larger than 0.09 SD. Finally, there is a 1.85 percentage point (0.09 SD) *decrease* in passing scores that remains statistically significant at the 1% level, and the 95% CI rules out any decrease smaller than 0.05 SD.

On the 8th Grade ELA Exam, when only considering fixed effects, there is a 1.55 percentage point (0.13 SD) decrease in Level 1 scores, a 1.25 percentage point (0.16 SD) increase in Level 4 scores, and a 2.2 percentage point (0.10 SD) increase in passing scores. As is the case with the 4th Grade ELA Exam, these scores are all significant at the 1% level. However, when we introduce school-specific linear time trends, we find that many of the same effects disappear; now, there is a 0.79 percentage point (0.07 SD) decrease in Level 1 scores, a 0.05 percentage point (0.01 SD) increase in Level 4 scores, and a 3.38 percentage point (0.16 SD) *decrease* in passing scores. The Level 1 and Level 4 score differences are not statistically significant at any reasonable level, and passing scores are significant at the 1% level. The 95% CI for Level 1 allows for decreases as large as 0.15 SD, the Level 4 CI rules out all effects (positive or negative) larger than 0.1 SD, and the passing score CI rules out any decrease smaller than 0.1 SD and allows for decreases of over 0.2 SD.

XII(c): Math Test Scores

When considering the 4th Grade State Math Assessment, we see results similar to those from the ELA exams: a 3.47 percentage point (0.38 SD) drop in Level 1 scores, a 9.58 percentage point (0.54 SD) increase in Level 4 scores, and an 11.1 percentage point (0.58 SD) increase in passing scores, all significant at the 1% level. However, the percentages change drastically when we include school-specific linear time trends, but not as much as in the ELA results. Here, there is a 2.58 percentage point (0.28 SD) drop in Level 1 scores, a 3.03 percentage point (0.17 SD) increase in Level 4 scores, and a 4.07 percentage point (0.21 SD) increase in passing scores. These results are also all significant at the 1% level. The Level 1 CI allows for drops as large as 0.37 SD and rules out any effect smaller than 0.19 SD, the Level 4 CI rules out any effect smaller than 0.12 SD and larger than 0.22 SD, and the passing score CI rules out any increase smaller than 0.16 SD and larger than 0.26 SD. The main difference between these results and the ELA results is that the Math results have smaller point estimates but the same direction of change when switching from the fixed effects model to the linear time trends model. In contrast, the ELA test scores change direction when using the time trends model.

Now, we consider the 8th Grade State Math Assessment. Using time and school fixed effects (Equation 2), NYC public schools see a 10.2 percentage point (0.55 SD) drop in Level 1 scores, a 3.82 percentage point (0.39 SD) increase in Level 4 scores, and a 13.4 percentage point (0.55 SD) increase in passing grades. The regression coefficients are all significant at the 1% level due to the large sample size (N=17,515). The relevant coefficients are outlined in Table 6. When school-specific linear time trends are brought under consideration our effects change somewhat drastically. There is still a 5.4 percentage point (0.30 SD) decrease in Level 1 scores, but now there is a 2.59 percentage point (0.27 SD) *decrease* in Level 4 scores and a 1.06 percentage point (0.04 SD) increase in passing grades. The first two effects are statistically significant at the 1% level, but the passing grades effect is not statistically significant at any level ($p>0.1$). The Level 1 CI rules out any effect smaller than 0.2 SD, and the Level 4 CI rules out any decrease smaller than 0.12 SD, while the passing score CI rules out most big effects (any increase larger than 0.1 SD). This development may mean that the effect of this policy on passing grades cannot be differentiated from the effect of including linear time trends.

XII(d): Other Behavioral Outcomes

Now, we move on to attendance records, the percentage of enrolled students attending class on an average day. This regression is split between high schools, middle schools, and elementary schools: the fixed effects regression demonstrates that there is a 0.15 percentage point (0.02 SD) decrease in high school attendance (not significant, $p>0.1$), a 0.97 percentage point (0.24 SD) increase in middle school attendance ($p<0.01$), and a 1.15 percentage point (0.34 SD) increase in elementary school attendance ($p<0.01$). These coefficients can be found in Table 7. However, when we include school-specific linear time trends, we see slightly different results. In this case, high school attendance rates go up 0.13 percentage points (0.02 SD, insignificant, $p>0.1$), middle school attendance rates increase by 0.37 percentage points (0.09 SD, $p<0.1$), and elementary school attendance rates increase by 0.64 percentage points (0.19 SD, $p<0.01$). The high school attendance rate effects are not large, ruling out any change in either direction of more than 0.1 SD. Middle school attendance is solidly positive, with the lower CI bound at zero while allowing effects as large as 0.18 SD. Elementary school attendance has large upside, allowing for effects as large as 0.26 SD while ruling out any changes smaller than 0.12 SD.

Now, we move to suspension rates. This measure is coded as suspensions per student in each school in each year, and the regressions are split into high school, middle school, and elementary school results. Overall, suspensions have generally increased in NYC schools since the treatment took place, in both the fixed effects and linear time trend models. When considering the fixed effects model, there is a 1.01 percentage point (0.12 SD) increase in suspensions per high school student, a 2.01 percentage point (0.21 SD) increase in suspensions per middle school student, and a 0.47 percentage point (0.07 SD) increase in suspensions per elementary school student. All these results are significant at the 1% level, and the coefficients can be found in Table 8. When we include linear time trends, suspensions per student increase at an even larger magnitude. In this model, there is a 3.95 percentage point (0.45 SD) increase in suspensions per high school student, a 4.12 percentage point (0.43 SD) increase in suspensions per middle school student, and a 1.61 percentage point (0.26 SD) increase in suspensions per elementary school student, all significant at the 1% level. The high school and middle school suspension CIs rule out any increase smaller than 0.35 SD, while the elementary school suspension CI rules out any increase smaller than 0.2 SD.

Next, this study considers aspirations, or post-high school plans. In these outcomes, linear time trends do not change the regression coefficients in a substantial manner, which means that these results are likely more steadfast than the fixed effects coefficients from other outcome variables. It is important to note that these aspirations do assume that the students in question will graduate, so these results leave out high school dropouts and only focus on the post-high school plans of graduates. The only outcome that changed substantially after introducing linear time trends was attendance at a two-year college: in the fixed effects model, there was a 9.65 percentage point (0.56 SD) increase in students attending two-year colleges after high school ($p<0.01$), while the linear time trends model only showed a 1.57 percentage point increase (0.09 SD, $p>0.1$). Four-year college attendance saw a 6.43 percentage point (0.29 SD) increase in the fixed effects model ($p<0.01$), and a 12.5 percentage point increase (0.56 SD) in the linear time trends model ($p<0.01$). The four-year college CI rules out any effect smaller than 0.4 SD with effects that could be as large as 0.7 SD, while the two-year college CI rules out any effect larger than 0.23 SD. Therefore, the effect on students planning to attend four-year colleges is large and significant.

Plans to enter the workforce immediately upon leaving high school saw a change in the sign of the point estimate, but the point estimates in both cases were small: there was an 0.85 percentage point (0.09 SD) decrease in immediate employment plans ($p<0.05$), and a 1.3 percentage point (0.14 SD) increase in the linear time trends model ($p<0.01$). However, standard errors for employment were large, so effects could be as high as 0.24 SD. There is a 0.23 percentage point (0.07 SD) decrease in military enlistment in the fixed effects model (insignificant, $p>0.1$), and a 0.24 percentage point (0.08 SD) decrease in the linear time trends model (insignificant, $p>0.1$), so either way, the effect is statistically insignificant and therefore cannot be statistically differentiated from zero in either case. The 95% CI suggests that a reduction in military plans could be as low as 0.17 SD while not increasing past 0.02 SD, so despite the statistical insignificance of this reading, military plans likely decreased in this sample.

Next, we consider the effect on a post-secondary institution not billed as a college, such as a trade school or vocational school; the fixed effects model shows a 0.40 percentage point (0.09 SD) increase in non-college post-secondary education (insignificant, p>0.1), and the linear time trends model shows a 0.25 percentage point increase (0.05 SD, p>0.1). These effects have a large standard error, so decreases could be as large as 0.15 SD while increases could be as high as 0.26 SD. The relevant fixed effects coefficients are found in Table 9, and the linear time trend coefficients are found in Table 10. These aspirations were by no means the only responses in this dataset, but they were by far the most common, allowing us to neglect the other responses.

Finally, we examine graduation rates, the number of students in each class who receive high school diplomas at the end of their senior year. The model choice only affects the magnitude of the results in this case, and does not affect the sign or the statistical significance of the coefficient. In the fixed effects model, we see a 7.38 percentage point (0.22 SD) increase in graduation rates for NYC schools, and in the linear time trends model, we see a 3.95 percentage point (0.12 SD) increase in graduation rates for those same NYC schools. The 95% CI for graduation rates rules out any increase larger than 0.16 SD. Both models are statistically significant at the 1% level, and the coefficients can be found in Table 11. Overall, it appears that the introduction of linear time trends only explained approximately half the total effect of this policy on graduation rates, which means that already-present improvement in NYC schools in the years before FSF was implemented only explains half of the total change in graduation rates.

XIII: Heterogeneity, Test Score Results by Race

This study also examines heterogeneity in outcome by different subgroups, such as race, gender, socioeconomic status, English proficiency, presence of disability, and so on. The purpose of this subgroup analysis is to determine if FSF has any significant positive effect on one or two specific groups so that future changes may be more targeted to groups that are left behind by current policy. This analysis also uses the fixed effects and linear time trends comparison for the same reason as the overall results: to determine whether schools would likely have improved anyway had FSF not been implemented.

This study will examine all four test scores split by race. First, this study looks at the results for black students. The relevant coefficients for this subgroup can be found in Tables 12 through 15. The expected decrease in Level 1 scores and increases in Level 4 and passing scores are present in the fixed effects model. However, once linear time trends are brought into effect, statistical significance entirely disappears in 8th grade math test scores and 4th grade ELA scores. In 8th grade ELA scores and 4th grade math scores, we see a statistically significant increase ($p<0.01$) in Level 1 scores upon the introduction of linear time trends. The 95% CIs for black students suggest a minimum increase of 0.34 SD in Level 1 scores, a maximum increase in Level 4 scores of 0.13 SD, and maximum increase in passing scores of 0.06 SD with a potential decrease as large as 0.14 SD. The effects in this circumstance are somewhat ambiguous, but results largely do not support the notion that FSF caused an increase in student performance.

Next, this study examines test scores for white students. These coefficients can be found in Tables 16 through 19. For white students, the expected decrease in Level 1 scores and increases in Level 4 and passing scores are still present. Once linear time trends are brought into the regression, statistical significance disappears in all measures for 8th grade math test scores. Along those same lines, for 4th grade math test scores, there is a statistically significant increase in Level

1 scores and a statistically significant increase in Level 4 scores, but passing scores remain statistically zero. Both 4th and 8th grade ELA scores see an exact reversal of all coefficient signs while still maintaining 1% statistical significance. There are increases in Level 1 scores and decreases in both Level 4 and passing scores once accounting for previous trends. Overall, despite the shift in sign, the point estimates are small, so linear time trends may not change the results all that much. However, standard errors are large enough that we cannot rule out large increases in 8th grade Math Level 4 scores or 4th grade Math Level 1 scores.

Next, we look at outcomes for Hispanic students. These coefficients can be found in Tables 20 through 23. Just like black and white students, Hispanic students have a statistically significant decrease in Level 1 scores and statistically significant increases in Level 4 scores and passing scores. Also, just like black and white students, those relationships break down when linear time trends are put into effect. All statistical significance in 8th grade math test scores dissipates, and 8th grade ELA test scores make the same complete reversal made by white students (increases in Level 1 scores and decreases in Level 4 and passing scores). All 4th grade scores (both Math and ELA) retain positive effects for Level 4 and passing scores, but adopt statistically significant increases in Level 1 scores as well, creating an ambiguous result for younger students. Unfortunately, the 95% CIs show that we can rule out large positive effects almost entirely across the board for Hispanic students; the only possibility for positive effects greater than 0.3 SD for Hispanic students is on the 4th grade Math assessment, in which the maximum Level 4 score increase is 0.47 SD and the maximum passing score increase is 0.33 SD.

The next outcomes to consider are those of Asian students. These coefficients are found in Tables 24 through 27. These effects are far less statistically significant across the board. For 4th grade ELA test scores, there is no statistically significant effect on Level 1 scores, and statistically significant ($p<0.01$) effects on Level 4 and passing scores. Once linear time trends are introduced, all significance disappears. For 8th grade ELA test scores, the only measure that has any reasonable level of statistical significance is a 2.34 percentage point (0.15 SD) increase in Level 4 scores using linear time trends ($p<0.1$). All other measures have $p>0.1$. Finally, the 4th and 8th grade math test scores in this cohort exhibit the usual trends seen in previous racial cohorts, a decrease in Level 1 scores and increases in Level 4 and passing scores. Consistent with the results observed in other racial groups, once linear time trends are instituted, most statistical significance goes away (passing scores remain as a statistically significant increase at the 5% level). What stands out about the confidence intervals for Asian students is that standard errors are large, so the lack of statistical significance found throughout this population does not necessarily mean that the result will be close to zero. For example, for passing scores on the 4th grade ELA exam, we cannot rule out decreases as large as 0.14 SD or increases as large as 0.18 SD. Similarly, for 8th grade Math test passing scores, we cannot rule out decreases as large as 0.21 SD or increases as large as 0.20 SD. This large standard error makes it difficult to conclude much about this population.

Finally, we consider Native American students. The relevant regression coefficients for Native American student test scores can be found in Tables 28 through 31. The population of this subgroup is substantially lower than the other four, so the results look far larger but are not as clear-cut. When considering the fixed effects model, general results show a highly statistically significant decrease in Level 1 scores in all four tests (4th grade ELA, 8th grade ELA, 4th grade Math, and 8th grade Math), increases in Level 4 ELA scores, decreases in Level 4 math scores, and an ambiguous effect on passing grades. However, once we include linear time trends, the statistical significance of many of these effects disappears. In most of these tests, there is still a decrease in Level 1 scores, but statistical significance is less common. Decreases in Level 4 scores become more common, with little statistical significance, and not a single passing score coefficient is statistically significant. For 4th grade ELA test scores, the Level 4 score decrease is so large that we can rule out any decrease smaller than 2.35 SD. For 8th grade ELA scores, the standard errors are so large that passing scores could drop by as much as 1.17 SD and increase by as much as 3.82 SD. 4th and 8th grade Math scores follow similar patterns to their ELA counterparts, with 4th grade scores showing large decreases and Math scores showing large standard errors that obscure most solid conclusions.

XIV: Results by Other Subgroups

The test score data is also split into four other dichotomous subgroups: socioeconomic status (SES, economically disadvantaged versus not economically disadvantaged), gender (male versus female), special education status (disabled versus general education), and English proficiency (English proficient versus not English proficient). These data define "economically disadvantaged" students as students whose families participate in some economic assistance program, be it TANF, EITC, SNAP, Medicaid, or some other social safety net program aimed to help poorer families.

First, we will consider test score results separated by SES. The relevant coefficients are found in Tables 32 through 39. The general trend found in these results reflects the previous literature, in that students whose families accept some form of public assistance perform at lower levels on these tests than do students from more financially well-off families. All four tests show a statistically significant increase in Level 1 scores ($p<0.01$) among economically disadvantaged students, even after linear time trends are put in place. Meanwhile, economically well-off students see a three to five-point decrease in Level 1 scores. Interestingly, economically disadvantaged students do see a statistically significant increase in Level 4 scores on both the 4th grade and 8th grade State Math Assessments ($p<0.01$), but statistically significant decreases in Level 4 scores on both the 4th grade and 8th grade ELA exams ($p<0.01$). For 8th grade ELA passing scores among high-SES students, we can rule out increases larger than 0.13 SD according to the 95% CI, and for 4th grade Math passing scores among low-SES students, we can rule out increases above 0.121 SD.

When splitting the population by gender, we come to the rather interesting conclusion that there is no discernable difference in effect between the male and female populations. Simply put, boys and girls are affected by this policy in basically the same way. There is a significantly negative effect on Level 1 math scores and a significantly positive effect on Level 4 math scores for 4th graders and 8th graders of both genders ($p<0.01$). There is a statistically significant decrease in passing ELA scores for both genders in both grades ($p<0.01$), and there is very little statistical significance in the Level 1 and Level 4 ELA score changes in either grade for either gender. Even the standard errors are essentially the same across the board, so confidence intervals can rule out the same effects. For example, both male and female results can rule out increases in 4th grade ELA Level 1 scores of more than 0.13 SD. The parallels between the two genders suggests that there is no gender gap in this policy, which makes it easier to generalize the effects of FSF to all students. The relevant coefficients are found in Tables 40 through 47.

A rather promising result emerges when test scores are segregated by special education status. The special education and general education coefficients are found in Tables 48 through 55. This study finds that, while the effects of FSF on the special education community (in percentage points) are not always positive in an absolute sense, they are better than the results for the general education community almost entirely across the board. The disparity in results tends to be much clearer in 4th grade Math scores than in any other test score. When the results are measured in standard deviations from the pre-treatment mean, the general education effect appears larger, most likely due to the fact that the pre-treatment mean of general education students scoring at Level 1 was already low (3.42%). Both special education and general education students experience a statistically significant ($p<0.01$) decrease in Level 1 scores, increase in Level 4 scores, and increase in passing scores; however, the decrease in Level 1 percentages is 3.82 percentage points (0.18 SD) for disabled students and 1.54 percentage points (0.27 SD) for general education students, which means the decrease more than doubled for disabled students. Level 4 scores went up by 3.08 percentage points (0.17 SD) for general education students, but went up by 4.13 percentage points (0.33 SD) for disabled students. Finally, the biggest disparity of them all: passing grades went up 2.86 percentage points (0.16 SD) for regular students and 11.6 percentage points (0.43 SD) for disabled students, more than a 300% increase over the general result. For 8th grade Math scores for disabled students, we know based on the 95% CI Level 1 scores decrease by a minimum of 0.385 SD

Finally, we examine the difference between students proficient in English and students not proficient in English. The effects can be found in Tables 56 through 63. The disparity of effects between these two groups is not as large as the gap between disabled and general education students, but it is still worth noting. English-proficient students see a statistically significant *increase* ($p<0.01$) in Level 1 4th grade ELA scores, but non-English proficient students do not see a statistically significant change in that same test. For English-proficient 4th grade ELA test scores, neither Level 4 scores and passing scores has a large effect in either direction according to their 95% CIs, so we can rule out Level 4 effects greater than 0.05 SD and passing score effects greater than 0.08 SD in either direction. The 8th grade ELA scores CI for non-English proficient students (Table 59) shows that, despite statistical insignificance, decreases in Level 1 scores could be as high as 0.42 SD, suggesting more upside to this policy than previously thought. Other than one clear example, the overall result is that English-proficient students experience negative effects because of FSF, and students who are learning English have a more positive reaction (but not necessarily positive in an absolute sense) to the change in funding structure brought about by FSF.

Graduation rates are also split into two subgroups: special education and general education. This split uncovers a rather unfortunate result: even after including linear time trends, disabled students experience a 12.4 percentage point (0.39 SD) decrease in graduation rates ($p<0.01$), while general education students see a 6.04 percentage point (0.18 SD) increase in the same metric ($p<0.01$). CIs shows that disabled graduation rates fall by at least 0.29 SD and general education graduation rates jump by at least 0.135 SD. This comes from the same population that saw special education students perform better on all four test score metrics mentioned in this study. The relevant coefficients can be found in Table 64.

XV: Discussion

This section will consider potential causes of the relationships outlined in the previous section. Overall, the introduction of linear time trends into our analysis appears to change the results rather substantially. When only using year and school fixed effects in the test score data, there is a general decrease in Level 1 state test scores, a general increase in Level 4 state test scores, and a general increase in passing grades. However, when we introduce linear time trends, those effects disappear, and we end up with either statistically insignificant effects with small point estimates or highly statistically significant negative effects on Level 4 scores and passing grades with small point estimates. This trend applies to all test scores in this study, which means that this policy most likely had no appreciable effect on test scores because all these test scores depended on pre-existing linear trends. There is a caveat: when effects phase in slowly, linear time trends make it a bit more difficult to identify the effect of a policy. However, it is likely that NYC schools were already improving, and test scores were already trending upward before FSF.

Another important takeaway from these results is that math test score improvements appear to be more resilient than ELA test scores, especially at the 4th grade level. It is possible that this could signify the importance of early intervention and funding fairness in a student's ability to grasp mathematical concepts. The lack of effect on overall test scores may indicate that this change in funding is too late of an intervention for reading comprehension skills. It may also indicate that funding structure does not actually have an effect on reading comprehension. These implications are not yet clear, and more research is needed to tease out these results.

Attendance rates have gotten better and suspension rates have increased for elementary, middle, and high schools since FSF was implemented, even when accounting for linear time trends. The improved attendance rates may show a positive effect that FSF has had on overall student performance, but the higher suspension rates could mean several potential implications. The higher suspension rates could mean that students are behaving worse, forcing schools to enforce more suspensions. However, it could also mean that schools are becoming stricter, leading administrations to crack down on previously permissible behavior. Further research would be needed to determine the reason behind the increased suspension rates.

The changes in aspirations since the implementation of FSF show that, even accounting for previous trends, four-year college enrollment plans increased by 12.4 percentage points (0.55 SD). This seems to indicate that more children who attend NYC public schools are planning to attend a four-year college than in years past, which bodes well for the effectiveness of this program. There was also a statistically significant increase in students who plan to enter the workforce immediately after high school, but the point estimate for this metric was small (1.3 percentage points, 0.14 SD) so it is not a cause for concern.

We can see which portions of the population experience the most success by splitting test scores by subgroup. Generally, white and Asian students experience larger improvements in test scores than do black and Hispanic students, a result that largely lines up with the previous literature. This may mean that changing the funding structure of schools is not enough to compensate for other disadvantages experienced by black and Hispanic NYC populations, and the solution may lie in other policy areas. This conjecture is reinforced when examining test scores broken down by SES; economically advantaged populations benefit far more from FSF than do economically disadvantaged populations on standardized tests. One would think that an increase of funds would help lower-income students, but the data says otherwise. This supports the notion that the solution for helping lower-income students lies in other policy areas besides school funding.

Test scores seem to change approximately the same for male and female populations, so FSF does not seem to favor one gender over the other. When splitting by special education status, we find one of the most promising statistics of the entire study; disabled students increase test scores by far higher margins than do general education students. This is most likely due to the FSF formula's high weight placed on a school's percentage of students in special education programs. A similar trend emerges for students not proficient in English; their 4th grade test scores in both English and math increase by more than those of students proficient in English. This result may suggest that reallocation of funds to ELL (English as a Learned Language) programs early in a child's education may make a difference in student performance, and the high weight placed on ELL programs by FSF may make that difference.

Unfortunately, even though disabled students benefit greatly in test scores, their success does not extend to graduation. Graduation rates increase for NYC public school students, but when the data is split by special education status, the success is limited almost entirely to general education students. Graduation rates for special education students plummet. It is possible that disabled populations may have increased over the years as more students qualify for classification. It is also possible that principals push more students to be classified as disabled in order to get more money for the school. Another potential reason is that a true effect on graduation rates may take a few years to filter through the school system, so FSF was too little too late for those in cohorts set to graduate soon after the program was implemented. Graduation requirements for disabled students could have also changed during this time period. None of these conjectures has empirical backing, but these are mere possibilities that warrant further research. Test score improvements are promising, but they do not mean much if they do not improve graduation rates.

Overall, many conflicting results of this study suggest that FSF is not unambiguously helpful. Test scores generally improved, but some of the most marginalized groups, namely economically disadvantaged, black, and Hispanic students all improved much less than richer, white, and Asian students. In some cases, marginalized communities did not improve at all.

Point estimates are mostly very small or statistically insignificant for this study, but in many cases, standard errors are also small. Therefore, even if an effect is statistically insignificant, we can still rule out certain effects of a certain size or find the minimum size of other effects with 95% certainty. The zeroes we find are true zeroes, not imprecisely estimated zeroes. For example, we were able to rule out almost all large improvements for Hispanic students despite statistical insignificance due to small standard errors. Either way, it does appear that FSF has had a positive overall impact on these populations, even if they perform relatively worse than their more fortunate counterparts. While the test scores of disabled students did improve, their graduation rates actually declined substantially, which may suggest that FSF does not hold the solution for this population. Instead, some other intervention may be necessary to improve the graduation rates of the disabled population.

XVI: Robustness Checks

XVI(a): Propensity Score Matching

After analyzing the primary research design, fixed effects OLS regressions, this study uses a different research design to determine if the results from the OLS regressions and linear time trends are resistant to changes in identification strategy. Here, we use matching by propensity scores; in this system, one NYC school and one non-NYC school would be matched based on propensity score, which is interpreted as the probability that a school is an NYC public school given its covariates. This new research design matches the NYC and non-NYC schools with the most similar propensity scores, and then measures their differences in dependent variables (educational attainment, graduation, college attendance, etc.).

A major weakness of this design is that propensity scores are based on the covariates of the year before the treatment date. This does not allow us to see the trend before the treatment, which prevents us from knowing if, without FSF, one group would have performed better than the other anyway. The coefficients of these results can be found for these results can be found in Tables 65-67. If we compare the propensity score matching results with the OLS results, we see that the matching coefficients are very similar to the results from the fixed effects OLS regressions (regression (2), without the linear time trends).

The results do not perfectly line up, but they have mostly the same signs and are at roughly the same order of magnitude. For example, in the fixed effects regression, aspirations of attending a four-year college increased by 6 percentage points (0.28 SD) while they increased by 11 percentage points in the matching model. In almost every test score result from the OLS fixed effects model, Level 1 scores decreased while Level 4 and passing scores increased. The propensity score matching results showed the same trend; in the OLS model, Level 1 scores on the 8th grade Math Assessment decrease by 10 percentage points (0.55 SD) while Level 4 scores increase by 3.8 percentage points (0.39 SD) and passing scores increase by 13.4 percentage points (0.54 SD). Meanwhile, in the propensity score model, Level 1 scores decrease by 4.6 percentage points, Level 4 scores increase by 4.3 percentage points, and passing scores increase by 8.8 percentage points. These results suggest that these data are fairly robust to this change in identification strategy.

XVI(b): DFL Reweighting

The next robustness check uses DiNardo, Fortin, and Lemieux (1996), reweighting in combination with the fixed effects regression. This method was used by Yagan (2015) to study the effect of dividend taxes on investment. This process begins by generating the same propensity scores used in the previous subsection. However, instead of matching treatment and comparison units based on those propensity scores and measuring the average difference between them over time, this method generates weights for the treatment and comparison groups. The purpose of this process is to weight the comparison group covariates upward or downward to look like the treatment group. Then, we run the linear time trend regressions on all outcome variables, once without reweighting and once with reweighting. As stated before, covariate balance is not necessary for a difference in differences model, but it can help make the parallel trends assumption more believable.

The coefficients of all relevant outcome variables, including test score levels, graduation rates, attendance rates, suspensions, and aspirations are remarkably resilient to the reweighting process. As seen in Table 68, the difference in differences coefficients of the standard linear time trends regression and the linear time trends regression are nearly identical, only differing in most cases by one tenth of one percentage point. For example, the associated coefficient for attending a four-year college is 0.12545, which means that city schools experienced a 12.545 percentage point larger increase in plans to attend a four-year college than non-city schools after 2008. The reweighted coefficient is 0.12496, a difference of almost one half of one percentage point. This dynamic is common throughout almost all outcomes. Interestingly, graduation rates show the largest change in response to this new research design. It is possible that graduation rates are a noisy measure; some schools define graduation based on the number of people who receive a diploma in any given year, and some schools may include people who were left back at some point, which may skew results. However, this is not a cause for concern because the coefficient only changes by one percentage point, which is only an 8% change (changing from 11 to 12 percentage points). Therefore, these results serve as another robustness check in favor of the original linear time trends regression used by this study.

XVII: Limitations

XVII(a): Data Limitations

Before the 2005-2006 school year, the ELA and State Math Assessments were only administered to children in grades four and eight. Starting in that school year, the test began to be given to all children in grades three thru eight, so there is a massive increase in data at that time. This study will need data from further back than 2005-2006, so the data points from grades outside of four and eight were excluded in order to maintain continuity.

Unfortunately, NYSED did not start recording cross-group data until 2012, which is too recent to use in this study, and NYSED does not make individual-level microdata publicly available, so we cannot examine cross-group terms. For example, we cannot examine how this policy affects black women, or economically disadvantaged white students, etc.; we can only measure how this policy affects black students as a whole, or economically disadvantaged people as a whole, and so on. This facet of the data may make the conclusions gleaned from these data less specific than they would otherwise be.

In addition, NYSED does not record all subgroups for all outcome variables. For example, test score data is split by race, SES, English proficiency, and special education status, so the test score outcome variables provide the greatest amount of subgroup detail out of all relevant outcome variables. However, graduation rates are only split by special education status, and attendance, suspensions, and aspirations are not split by subgroup at all. It would have greatly benefitted this study if we were able to look at graduation rates by race, or aspirations by SES, or any of the other breakdowns made available to this study in the test score data.

Definitive conclusions about FSF would likely benefit from its being implemented in a different city, a city that has been collecting and standardizing its data for enough years prior to treatment that the actual treatment effects can be isolated and studied for different subgroups and cross-groups over time. Although it is unlikely that a city would make such a decision solely to aid the ability to determine the effectiveness of FSF, this would be a clear way to see if FSF does work. However, until that day comes, these results are all we have on which to base our policies.

XVII(b): Research Design Limitations

Some outcome variables do not exhibit parallel trends before the treatment date, and therefore, determining causal effects with regression coefficients is far more difficult. 4th grade

Math scores and 8th grade Level 1 Math scores are some examples of non-parallel pre-treatment trends according to Figures 1 thru 12.

In addition to that, there may be confounding factors that come into play here. For example, right after the treatment year (2007-2008), the global financial crisis and the Great Recession caused a great deal of financial turbulence across the entire country, but hit NYC especially hard due to its heavy presence of financial sector jobs. Another possibility is that the linear time trend results suggest that NYC versus non-NYC schools may not be a perfect comparison group. This study does not separate out any effect that the recession may have had on school finances or budgets, which may pose a problem for making any causal claims about any results found in this study.

One example of a major policy change after the implementation of FSF that did *not* cause bias in these results is Common Core. Based on trends in Figures 1-12, there is a sharp break in the data right before the 2011-2012 school year, most likely as a result of the implementation of new Common Core testing standards. Since the change occurred across the entire state, the levels of test scores might have changed, but the difference between NYC and non-NYC schools did not change. To make sure Common Core did not change any results, the data was rerun with the 2012-2015 entries removed, and the fixed effects and linear time trends regressions were run again. There was very little change in regression coefficients, so Common Core did not change our results.

XVIII: Policy Implications

The next step is to determine what these conclusions mean for public policymakers. As stated at the beginning of this study, a lack of statistically significant effects likely means one of three things: FSF does not actually work, FSF has not been in place for long enough to see its true effects, or the redistribution put in place by FSF is not large enough to generate any observable change. Another issue here is that there are many confounding factors at play such as changing tests, changing curricula, the 2008 financial crisis and subsequent recession, and so on. Because of this, we would most likely need more data and a cleaner treatment without as many confounding effects in order to determine if FSF truly works. If it so happens that the policy intervention was too small, and we need to spend more money in each school to see positive results, we begin to run into an efficiency problem. The NYC DOE would need to determine how much money it is willing to spend per percentage point increase in student performance.

If we compare these results to other selected policy interventions, we can see that the changes in test scores do compare closely to some direct incentive-based structures. Allan and Fryer (2011) and Fryer (2013) show the effects of paying students for increases in performance, and the results shown in these studies amount to changes in 4th grade test scores of less than one tenth of one standard deviation without statistical significance ($p>0.1$). Meanwhile, FSF shows a statistically significant ($p<0.01$) decrease in 4th grade ELA test scores and a statistically significant ($p<0.01$) increase in 4th grade math test scores. These results directly conflict with each other, so it is difficult to conclude that FSF has a better effect than a direct incentive regime or any other potential reform.

Some of these results do indicate that the funding change does help disabled students score higher on standardized tests, so it may be advisable to increase funding to schools with higher populations of disabled students, a major part of the FSF policy change. However, the effects on other populations do not appear to be large or definitive enough to declare that this policy works for sure. Given the differences between the coefficients in the fixed effects model and the coefficients in the linear time trends model, it is possible that NYC schools were improving anyway, even before the implementation of FSF.

The lack of definitive effect is likely a result of the fact that changes in funding may not be able to fix failing schools by themselves. Once the funding structure changes, other policy changes dealing with teacher quality, school infrastructure and technology, and curriculum may need to be tested to start seeing real positive effects. Schools can begin to work on their other problems once they have the money to do so. Based on this notion, it would be advisable to leave FSF in place while focusing on other issues that can now be addressed.

XIX: Future Research

One future direction in which this line of study could be taken involves the difference between rural and urban schools. Obviously, there are no rural schools in New York City, but there are urban areas in the rest of New York State such as Buffalo, Albany, Rochester, Binghamton, and Syracuse. These areas exhibit characteristics much more similar to those of New York City public schools, in that they most likely have more minority representation and a greater percentage of students with limited English proficiency, and far less like their rural counterparts, which are demographically more white and homogeneous. These data could be approximately split into urban and rural schools using ZIP codes, which would allow us to control for the differences between urban and rural schools.

Another possible avenue for future research includes the formulation of FSF itself. This study has largely treated FSF as a monolith, a black box of a policy that either stays in place in its entirety or does not exist. However, FSF can be changed, so it is possible that a different formula of funding redistribution may be better-suited for New York City. A future study could subject the FSF formula to various stress tests to determine how funding allocation would change as a result. This sort of study could give rise to a randomized experiment; if we were to randomize which schools received the funding change, we could compare student performance changes within New York City among similar schools. This would provide far better covariate balance between the populations of constituent schools over the life of the study.

Another realm this study does not address is the impact of FSF on charter schools. This data set does designate charter schools, so it would be possible to perform that analysis. It is not clear exactly how funding differs in charter schools compared to regular public schools, but after considering the extra autonomy given to charter schools to run themselves how they see fit, there may be more variation in how extra FSF money is spent. Despite the increased variation in methods and teaching styles of charter schools, these schools still provide standardized test score and graduation data, so they can be easily integrated into a research design such as this.

As discussed in the limitations section, this study would have been far more robust if subgroup and cross-term data had existed before 2007-2008. A future study might be able to try implementing FSF in a different city with better pre-existing data. For example, if Chicago, a large city with a school system in need of reform, has subgroup and cross-term data going back long enough to establish previous linear time trends, a switch to FSF could be studied in far greater detail. It might not be able to say much on the effectiveness of FSF in New York City, but it may speak on the effectiveness of funding restructures in general. It may provide greater insight into whether FSF and other policies like it should be implemented in other cities across the country.

XX: Conclusion

There is no dispute that there is still a high level of inequality between children in New York public schools. The main goal of Fair Student Funding is to start making steps toward reducing that inequality so that all residents of New York City can obtain access to the highest quality education possible and close the gap between advantaged and disadvantaged communities across New York City.

This study has served as an initial foray into the currently untapped field of student performance in response to public funding structure changes. While there are some limitations within the data and research design of this study due to external factors, it lays the groundwork for potential policy action and future public school funding research. Overall, these data show that Fair Student Funding has had a generally positive impact on test scores, attendance records, and graduation rates for most subgroups across New York City's public schools. Some improvements were not statistically significant, and some subgroups even declined in performance. This study agrees with the previous literature in its conclusion that poor, black, and Hispanic populations generally do not improve as well as other subgroups in response to educational policy changes. The improvements in test scores are promising, and may suggest that FSF should stay in place. However, the effect on the graduation rates of disabled students raises some concerns, and may require additional policy intervention.

There is no silver bullet to solve the problems in New York City's public education system. Inadequate funding is not the only problem faced by the large majority of New York City's schools, and fixing that problem will not automatically make all schools better in the span of less than a decade. However, FSF does open the door for the NYC DOE to try other policy options in schools that need the most improvement. The idea of a redistributive funding structure cannot solve all the problems of the city's school system, but it is a step in the right direction.

XXI: References

Allan, B. M., & Fryer, R. G. (2011). The Power and Pitfalls of Education Incentives. The Hamilton Project, The Brookings Institution.

Baicker, K., Finkelstein, A., Song, J., & Taubman, S. (2013). The Impact of Medicaid on Labor

Force Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment. The New England Journal of Medicine, 368(18).
Baker, B. D. (2009). Within-district resource allocation and the marginal costs of providing equal educational opportunity: Evidence from Texas and Ohio. Education Policy Analysis Archives, 17(3). February 2009.
Baker, B. D. (2014, January 24). School Funding Fairness in New York State: An Update for 2013-14. Alliance for Quality Education.
Barnett, W. Steven. Long-Term Cognitive and Academic Effects of Early Childhood Education on Children in Poverty. Preventive Medicine, March 1998, Vol.27(2), pp.204-207
Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences -In-Differences Estimates? The Quarterly Journal of Economics, 119(1), 249-275.
Bettinger, E., & Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. Journal of Public Economics, 90(8-9), 1625-1648.
Buck, R., & Deutsch, J. (2014). Effects of poverty on education. International Journal of Human Sciences, 11(2), 1139.
Chakrabarti, R., & Setren, E. (2011, December). The Impact of the Great Recession on School District Finances: Evidence from New York. SSRN Electronic Journal.
Chetty, R., Hendren, N., & Katz, L. (2015). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. American Economic Review, 106(4), 855-902.
Contributing Factors: Disparities In 2005 Classroom Spending. (2007, October). Retrieved December 6, 2016, from http://www.ibo.nyc.ny.us/iboreports/FairStudentFunding1.pdf
Cooper, C. E., Crosnoe, R., Suizzo, M., & Pituch, K. A. (2009). Poverty, Race, and Parental Involvement During the Transition to Elementary School. Journal of Family Issues, 31(7), 859-883.
Dermott, E., & Pomati, M. (2015). 'Good Parenting Practices: How Important are Poverty, Education and Time Pressure? Sociology, 50(1), 125-142.
Dinardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica, 64(5), 1001-1044.
Escue, Carlee Poston. Adequate Yearly Progress as a Means of Funding Public Elementary and Secondary Education for Impoverished Students: Florida Funding. Journal of Education Finance, 2012, Vol.37(4), p.347-373
Fair Student Funding: Budgets That Put Students First. (2007, June). Retrieved December 6, 2016, from http://schools.nyc.gov/offices/d_chanc_oper/budget/dbor/allocationmemo/fy15_16 /FY16_PDF/FSF_Guide.pdf
Fryer, R. G. (2013). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. Journal of Labor Economics, 31(2), 373-407.
Graduation rate. (2013, June). Retrieved December 06, 2016, from http://schools.newsday.com/long-island/graduates/gradrate/
Imbens, G., & Rubin, D. B. (2015). Causal inference for statistics, social, and biomedical sciences: an introduction. New York: Cambridge University Press.
Jacob, B. A., & Ludwig, J. (2012). The Effects of Housing Assistance on Labor Supply: Evidence from a Voucher Lottery. American Economic Review, 102(1), 272-304.
Klein, J. I. (2015). Lessons of hope: how to fix our schools. New York: Harper.
Ladd, H. F. (2012). Education and Poverty: Confronting the Evidence. Journal of Policy Analysis

and Management, 31(2), 203-227.
New Funding Formula Seeks to Alter School Budget Disparities. (2007, October). Retrieved December 6, 2016, from http://www.ibo.nyc.ny.us/iboreports/FairStudentFunding2.pdf
New York State Education Department. (2015). 2002-2015 Report Card Database [Data file and code book]. Retrieved from https://data.nysed.gov/downloads.php
Rebell, Michael A. CFE v. State of New York: Past, Present, and Future. NYSBA Government, Law, and Policy Journal. Summer 2011, Vol. 13(1)
Schwartz, A.E.; Mccabe, B.J.; Gould Ellen, I.; Chellman, C.C. Public Schools, Public Housing: The Education of Children Living in Public Housing. Urban Affairs Review, 2010, Vol.46(1), pp.68-89
Tilak, Jandhyala B. G. (2002) Education and Poverty, Journal of Human Development, 3:2, 191-207,
UNICEF. (2000). Defining Quality in Education. New York: United Nations Children's Fund.
Yagan, D. (2015). Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut. American Economic Review, 105(12), 3531-3563.

XXII: Tables and Figures

Figure 1: Overall 4th Grade ELA Level 1 Annual Means

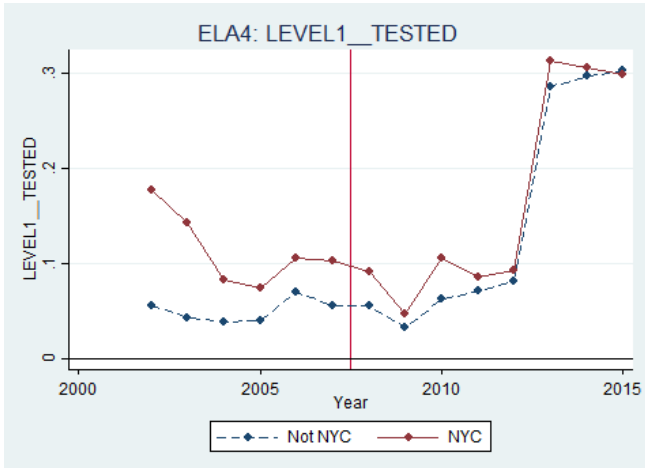


Figure 2: Overall 4th Grade ELA Level 4 Annual Means

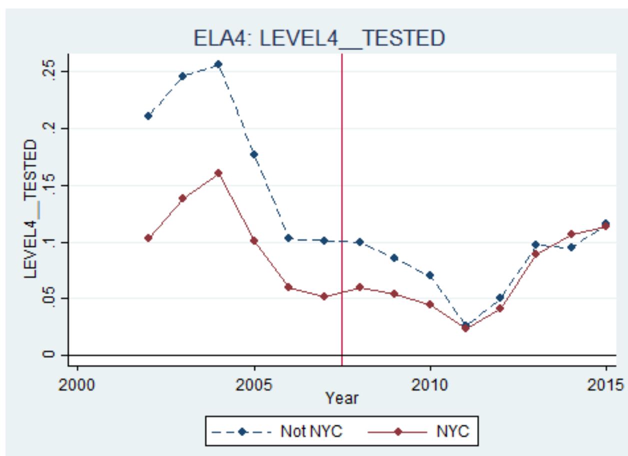
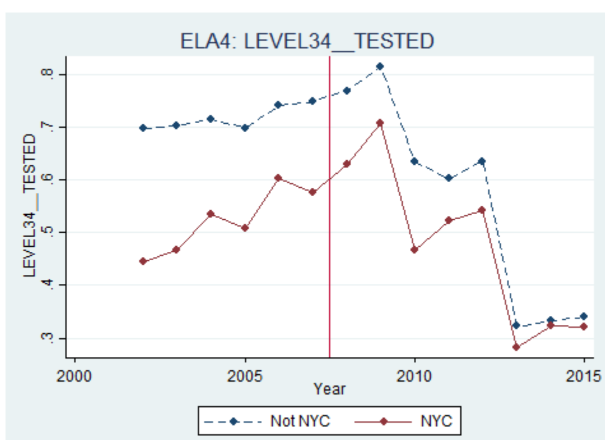


Figure 3: Overall 4th Grade ELA Passing Annual Means



Note: Treatment date is 2007-2008 school year. Graphs meant to show validity of parallel trends assumption.

Figure 4: Overall 8th Grade ELA Level 1 Annual Means

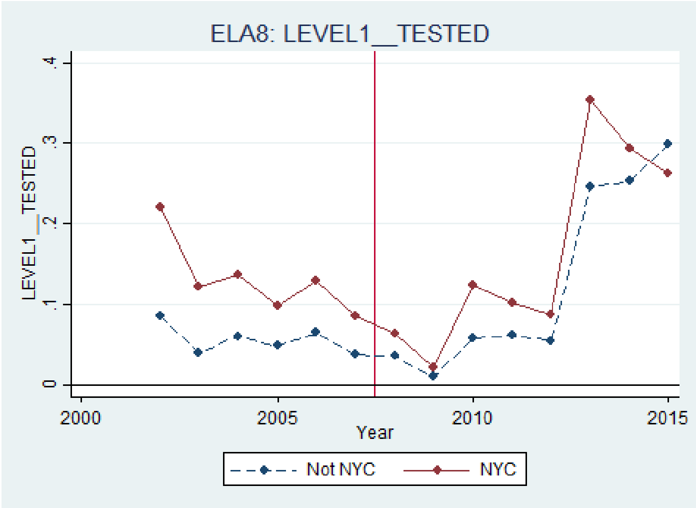


Figure 5: Overall 8th Grade ELA Level 4 Annual Means

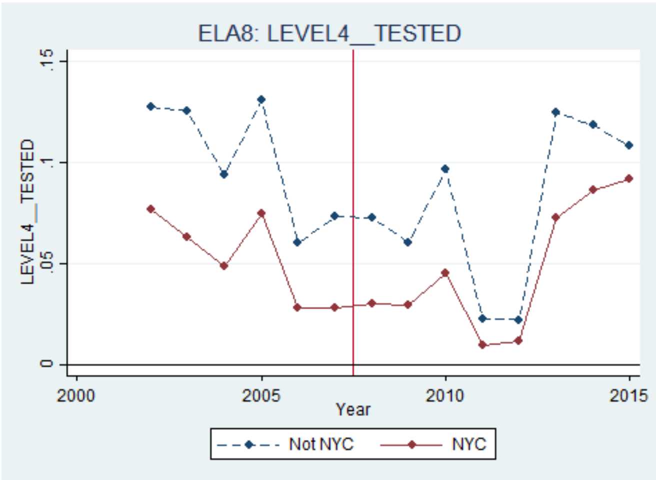
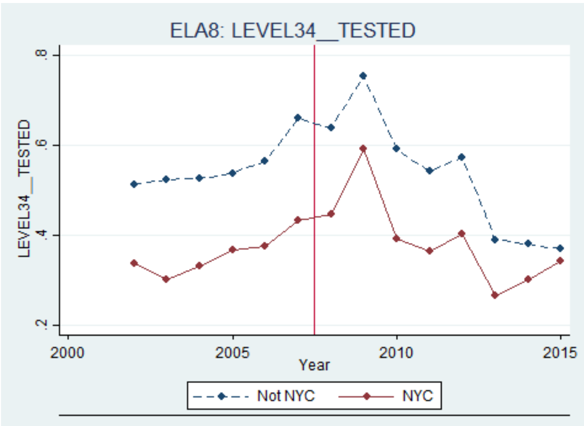


Figure 6: Overall 5th Grade ELA Passing Annual Means



Note: Treatment date is 2007-2008 school year. Graphs meant to show validity of parallel trends assumption.

Figure 7: Overall 4th Grade Math Level 1 Annual Means

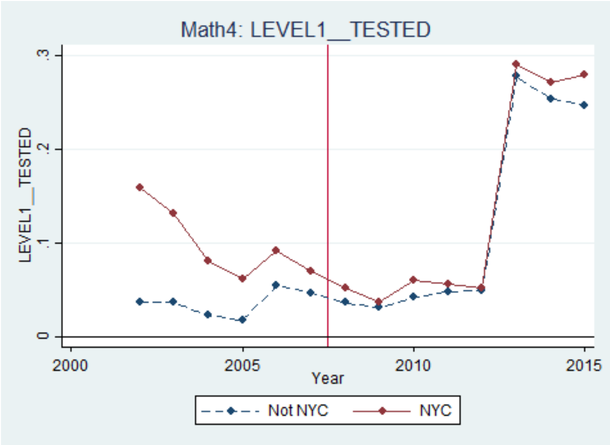


Figure 8: Overall 4th Grade Math Level 4 Annual Means

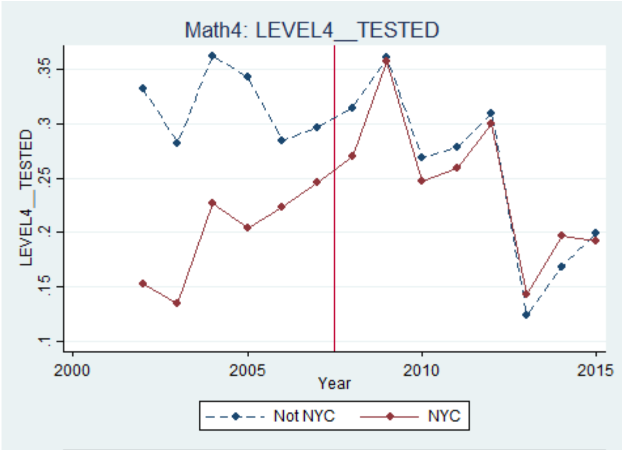
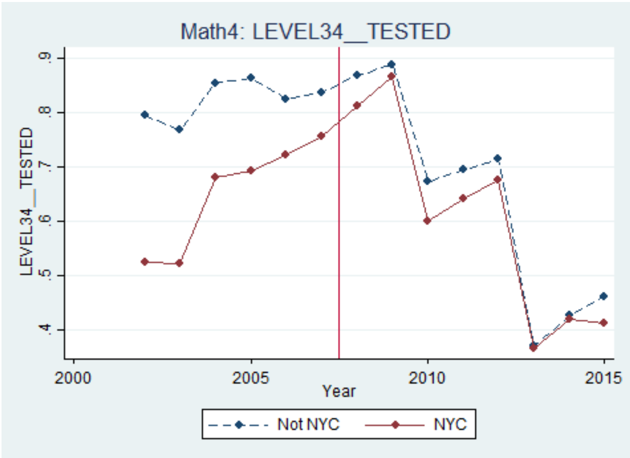


Figure 9: Overall 4th Grade Math Passing Annual Means



Note: Treatment date is 2007-2008 school year. Graphs meant to show validity of parallel trends assumption.

Figure 10: Overall 8th Grade Math Level 1 Annual Means

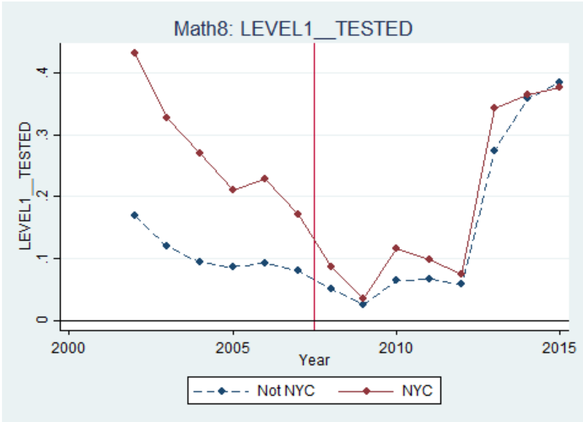


Figure 11: Overall 8th Grade Math Level 4 Annual Means

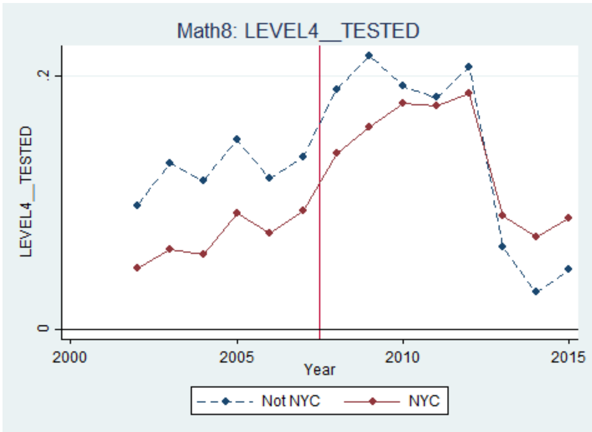
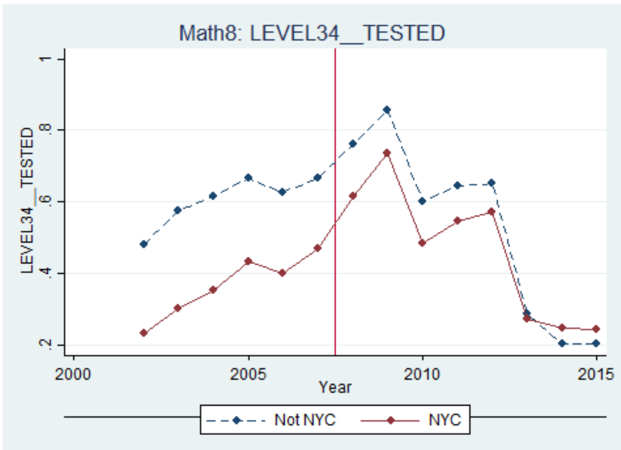


Figure 12: Overall 8th Grade Math Passing Annual Means



Note: Treatment date is 2007-2008 school year. Graphs meant to show validity of parallel trends assumption.

Table 1: Demographic Summary Statistics

	NYC (1)	Non-NYC (2)
Percent Free Lunch	75.99	23.76
	(410.1)	(21.73)
Percent Reduced Lunch	9.976	7.848
	(42.36)	(5.471)
Percent in Same School in Previous Year	85.70	93.42
	(18.32)	(12.02)
Percent with Limited English Proficiency	20.22	3.553
	(277.2)	(7.817)
Percent Native American	0.411	0.664
	(0.604)	(4.749)
Percent Black	37.23	11.79
	(28.96)	(19.86)
Percent Hispanic	40.06	8.969
	(25.50)	(13.69)
Percent Asian	10.13	3.080
	(15.75)	(4.615)
Percent White	12.08	74.92
	(19.21)	(28.84)
Number of Students	668.8	570.5
	(589.9)	(388.2)
Observations	1485	3064

mean coefficients; sd in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Summary Statistics for Pre-Treatment Outcome Variables (Fraction)

	NYC Pre-2008 (1)	Non-NYC Pre-2008 (2)	NYC Post-2008 (3)	Non-NYC Post-2008 (4)
Two-Year College	0.151	0.350	0.284	0.380
	(0.147)	(0.149)	(0.205)	(0.153)
Four-Year College	0.367	0.463	0.382	0.434
	(0.286)	(0.196)	(0.270)	(0.205)
Employment	0.0259	0.0983	0.0288	0.0988
	(0.0651)	(0.0900)	(0.0538)	(0.0876)
Trade School/Other	0.0171	0.0162	0.0180	0.0167
	(0.0648)	(0.0314)	(0.0408)	(0.0292)
Military	0.0113	0.0298	0.0102	0.0278
	(0.0225)	(0.0325)	(0.0184)	(0.0315)
Attendance Rate	0.895	0.949	0.903	0.947
	(0.0619)	(0.0292)	(0.0634)	(0.0423)
Suspension Rate	0.0352	0.0494	0.0396	0.0460
	(0.0516)	(0.0884)	(0.0606)	(0.0830)
Graduation Rate	0.389	0.794	0.589	0.815
	(0.348)	(0.238)	(0.352)	(0.267)
ELA4, Level 1	0.120	0.0554	0.177	0.151
	(0.113)	(0.0769)	(0.167)	(0.169)
ELA4, Level 4	0.101	0.173	0.0632	0.0748
	(0.112)	(0.131)	(0.0872)	(0.0746)
ELA4, Passing	0.520	0.703	0.458	0.548
	(0.203)	(0.172)	(0.240)	(0.251)
ELA8, Level 1	0.138	0.0701	0.179	0.147
	(0.148)	(0.101)	(0.185)	(0.178)
ELA8, Level 4	0.0475	0.0896	0.0395	0.0667
	(0.0782)	(0.0762)	(0.0773)	(0.0705)
ELA8, Passing	0.359	0.517	0.358	0.489
	(0.224)	(0.190)	(0.234)	(0.222)
Math4, Level 1	0.104	0.0397	0.150	0.124
	(0.113)	(0.0709)	(0.177)	(0.163)
Math4, Level 4	0.198	0.304	0.231	0.242
	(0.163)	(0.175)	(0.191)	(0.168)
Math4, Passing	0.647	0.813	0.577	0.631
	(0.206)	(0.158)	(0.270)	(0.256)
Math8, Level 1	0.273	0.124	0.209	0.179
	(0.214)	(0.150)	(0.222)	(0.219)
Math8, Level 4	0.0650	0.105	0.112	0.114
	(0.0975)	(0.0955)	(0.146)	(0.125)
Math8, Passing	0.369	0.571	0.429	0.487
	(0.243)	(0.221)	(0.287)	(0.308)
Observations	7909	18773	13110	23927

mean coefficients; sd in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: 4th Grade ELA Test Scores, Overall

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0305***	0.0618***	0.0913***	0.00230	0.00637**	-0.0185***
	(0.00272)	(0.00265)	(0.00352)	(0.00362)	(0.00277)	(0.00441)
did (in SD)	-0.323***	0.476***	0.456***	0.024	0.049**	-0.092***
	(0.029)	(0.020)	(0.018)	(0.038)	(0.021)	(0.022)
95% CI	(-0.036,-0.025)	(0.057,0.067)	(0.084,0.098)	(-0.005,0.009)	(0.001,0.012)	(-0.027,-0.01)
95% CI (in SD)	(-0.38,-0.266)	(0.437,0.515)	(0.421,0.491)	(-0.05,0.098)	(0.008,0.09)	(-0.135,-0.049)
Observations	32,548	32,548	32,548	32,548	32,548	32,548
R-squared	0.778	0.751	0.886	0.860	0.827	0.916
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0749	.1516	.6473	.0749	.1516	.6473
Pre-2007 SD	.0943	.1298	.2004	.0943	.1298	.2004
Schools	2808	2808	2808	2808	2808	2808

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 4: 8th Grade ELA Test Scores, Overall

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0155***	0.0125***	0.0220***	-0.00788	-0.000468	-0.0338***
	(0.00376)	(0.00275)	(0.00576)	(0.00547)	(0.00342)	(0.00592)
did (in SD)	-0.128***	0.158***	0.103***	-0.065	-0.006	-0.159***
	(0.031)	(0.035)	(0.027)	(0.045)	(0.043)	(0.028)
95% CI	(-0.023,-0.008)	(0.007,0.018)	(0.011,0.033)	(-0.019,0.003)	(-0.007,0.006)	(-0.045,-0.022)
95% CI (in SD)	(-0.189,-0.067)	(0.089,0.227)	(0.05,0.156)	(-0.153,0.023)	(-0.09,0.078)	(-0.214,-0.104)
Observations	17,600	17,600	17,600	17,600	17,600	17,600
R-squared	0.803	0.750	0.910	0.873	0.803	0.934
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0896	.0774	.4714	.0896	.0774	.4714
Pre-2007 SD	.1207	.0791	.2131	.1207	.0791	.2131
Schools	1866	1866	1866	1866	1866	1866

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 5: 4th Grade Math Test Scores, Overall

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0347***	0.0958***	0.111***	-0.0258***	0.0303***	0.0407***
	(0.00282)	(0.00349)	(0.00407)	(0.00416)	(0.00465)	(0.00472)
did (in SD)	-0.382***	0.539***	0.584***	-0.284***	0.170***	0.214***
	(0.031)	(0.020)	(0.021)	(0.046)	(0.026)	(0.025)
95% CI	(-0.04,-0.029)	(0.089,0.103)	(0.103,0.119)	(-0.034,-0.018)	(0.021,0.039)	(0.031,0.05)
95% CI (in SD)	(-0.443,-0.321)	(0.5,0.578)	(0.543,0.625)	(-0.374,-0.194)	(0.119,0.221)	(0.165,0.263)
Observations	32,560	32,560	32,560	32,560	32,560	32,560
R-squared	0.733	0.797	0.861	0.832	0.846	0.901
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0592	.2717	.7632	.0592	.2717	.7632
Pre-2007 SD	.0909	.1779	.1902	.0909	.1779	.1902
Schools	2812	2812	2812	2812	2812	2812

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 6: 8th Grade Math Test Scores, Overall

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.102***	0.0382***	0.134***	-0.0544***	-0.0259***	0.0106
	(0.00499)	(0.00529)	(0.00682)	(0.00885)	(0.00689)	(0.00825)
did (in SD)	-0.555***	0.391***	0.547***	-0.296***	-0.265***	0.043
	(0.027)	(0.054)	(0.028)	(0.048)	(0.071)	(0.034)
95% CI	(-0.112,-0.092)	(0.028,0.049)	(0.121,0.147)	(-0.072,-0.037)	(-0.039,-0.012)	(-0.006,0.027)
95% CI (in SD)	(-0.608,-0.502)	(0.285,0.497)	(0.492,0.602)	(-0.39,-0.202)	(-0.404,-0.126)	(-0.024,0.11)
Observations	17,515	17,515	17,515	17,515	17,515	17,515
R-squared	0.820	0.809	0.896	0.870	0.855	0.926
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1668	.0934	.5124	.1668	.0934	.5124
Pre-2007 SD	.1839	.0977	.2450	.1839	.0977	.2450
Schools	1864	1864	1864	1864	1864	1864

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 7: Overall Attendance Rates

VARIABLES	(1) Attendance (high)	(2) Attendance (middle)	(3) Attendance (elementary)	(4) Attendance (high)	(5) Attendance (middle)	(6) Attendance (elementary)
did	-0.00146	0.00970***	0.0115***	0.00126	0.00373*	0.00644***
	(0.00274)	(0.00140)	(0.000831)	(0.00335)	(0.00192)	(0.00114)
did (in SD)	-0.019	0.240***	0.343***	0.017	0.092*	0.192***
	(0.037)	(0.035)	(0.025)	(0.045)	(0.047)	(0.034)
95% CI	(-0.007,0.004)	(0.007,0.012)	(0.01,0.013)	(-0.005,0.008)	(0,0.007)	(0.004,0.009)
95% CI (in SD)	(-0.092,0.054)	(0.171,0.309)	(0.294,0.392)	(-0.071,0.105)	(0,0.184)	(0.125,0.259)
Observations	13,575	20,224	30,849	13,575	20,224	30,849
R-squared	0.851	0.602	0.524	0.901	0.680	0.611
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.9061	.9358	. 9423	.9061	.9358	. 9423
Pre-2007 SD	.0749	.0405	.0335	.0749	.0405	.0335
Schools	1705	2521	3240	1705	2521	3240

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are average daily school attendance rates, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 8: Overall Suspension Rates

VARIABLES	(1) Suspensions (high)	(2) Suspensions (middle)	(3) Suspensions (elementary)	(4) Suspensions (high)	(5) Suspensions (middle)	(6) Suspensions (elementary)
did	0.0101***	0.0201***	0.00469***	0.0395***	0.0412***	0.0161***
	(0.00378)	(0.00274)	(0.00136)	(0.00443)	(0.00360)	(0.00161)
did (in SD)	0.115***	0.209***	0.074***	0.450***	0.429***	0.256***
	(0.043)	(0.029)	(0.022)	(0.051)	(0.038)	(0.026)
95% CI	(0.003,0.018)	(0.015,0.025)	(0.002,0.007)	(0.031,0.048)	(0.034,0.048)	(0.013,0.019)
95% CI (in SD)	(0.031,0.199)	(0.152,0.266)	(0.031,0.117)	(0.35,0.55)	(0.355,0.503)	(0.205,0.307)
Observations	12,485	18,454	28,042	12,485	18,454	28,042
R-squared	0.621	0.625	0.603	0.751	0.751	0.735
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0825	.0625	.0230	.0825	.0625	.0230
Pre-2007 SD	.0877	.0960	.0630	.0877	.0960	.0630
Schools	1648	2430	3195	1648	2430	3195

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables suspensions per student, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 9: Time Fixed Effects and School Fixed Effects, Post-High School Aspirations

VARIABLES	(1) PER_2YR_ COLLEGE	(2) PER_4YR_ COLLEGE	(3) PER_ EMPLOYMENT	(4) PER_POST_ SECONDARY	(5) PER_ MILITARY
did	0.0965***	0.0643***	-0.00845**	0.00395	-0.00229**
	(0.00895)	(0.0136)	(0.00329)	(0.00264)	(0.000995)
did (in SD)	0.563***	0.286***	-0.094**	0.094	-0.073**
	(0.052)	(0.061)	(0.037)	(0.063)	(0.032)
95% CI	(0.079,0.114)	(0.038,0.091)	(-0.015,-0.002)	(-0.001,0.009)	(-0.004,0)
95% CI (in SD)	(0.461,0.665)	(0.166,0.406)	(-0.167,-0.021)	(-0.029,0.217)	(-0.136,-0.01)
Observations	16,813	16,813	16,813	16,813	16,813
R-squared	0.658	0.717	0.585	0.444	0.449
Year Fixed Effects	X	X	X	X	X
School Fixed Effects	X	X	X	X	X
Time trends					
Pre-2007 Mean	.3018	.4396	.0808	.0165	.0253
Pre-2007 SD	.1715	.2245	.0901	.0420	.0314
Schools	1710	1710	1710	1710	1710

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year and school fixed effects.

Table 10: Time F.E., School F.E., and Linear Trends, Post-High School Aspirations

VARIABLES	(1) PER_2YR_ COLLEGE	(2) PER_4YR_ COLLEGE	(3) PER_ EMPLOYMENT	(4) PER_POST_ SECONDARY	(5) PER_ MILITARY
did	0.0157	0.125***	0.0130***	0.00245	-0.00236
	(0.0122)	(0.0166)	(0.00464)	(0.00447)	(0.00158)
did (in SD)	0.092	0.557***	0.144***	0.058	-0.075
	(0.071)	(0.074)	(0.051)	(0.106)	(0.050)
95% CI	(-0.008,0.04)	(0.092,0.158)	(0.004,0.022)	(-0.006,0.011)	(-0.005,0.001)
95% CI (in SD)	(-0.047,0.231)	(0.412,0.702)	(0.044,0.244)	(-0.15,0.266)	(-0.173,0.023)
Observations	16,813	16,813	16,813	16,813	16,813
R-squared	0.744	0.792	0.669	0.578	0.521
Year Fixed Effects	X	X	X	X	X
School Fixed Effects	X	X	X	X	X
Time trends	X	X	X	X	X
Pre-2007 Mean	.3018	.4396	.0808	.0165	.0253
Pre-2007 SD	.1715	.2245	.0901	.0420	.0314
Schools	1710	1710	1710	1710	1710

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 11: Overall Graduation Rates

VARIABLES	(1) Graduation Rate	(2) Graduation Rate
did	0.0738***	0.0395***
	(0.00760)	(0.00748)
did (in SD)	0.225***	0.120***
	(0.023)	(0.023)
95% CI	(0.059,0.089)	(0.025,0.054)
95% CI (in SD)	(0.18,0.27)	(0.075,0.165)
Observations	14,346	14,346
R-squared	0.876	0.910
Year Fixed Effects	X	X
School Fixed Effects	X	X
Time trends		X
Pre-2007 Mean	.6803	.6803
Pre-2007 SD	.3287	.3287
Schools	2165	2165

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are school-level graduation rates, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 12: 4th Grade ELA Test Scores, Black

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0698***	0.0202***	0.0668***	0.0593***	0.00359	-0.00817
	(0.00615)	(0.00457)	(0.00823)	(0.00820)	(0.00452)	(0.0109)
did (in SD)	-0.561***	0.211***	0.328***	0.476***	0.038	-0.040
	(0.049)	(0.048)	(0.040)	(0.066)	(0.047)	(0.053)
95% CI	(-0.082,-0.058)	(0.011,0.029)	(0.051,0.083)	(0.043,0.075)	(-0.005,0.012)	(-0.03,0.013)
95% CI (in SD)	(-0.657,-0.465)	(0.117,0.305)	(0.25,0.406)	(0.347,0.605)	(-0.054,0.13)	(-0.144,0.064)
Observations	11,694	11,694	11,694	11,694	11,694	11,694
R-squared	0.761	0.557	0.777	0.844	0.685	0.832
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1276	.0727	.4849	.1276	.0727	.4849
Pre-2007 SD	.1245	.0957	.2038	.1245	.0957	.2038
Schools	1689	1689	1689	1689	1689	1689

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 13: 8th Grade ELA Test Scores, Black

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0319***	-0.00564*	-0.0143	0.0442***	-0.00459*	-0.0313***
	(0.00708)	(0.00301)	(0.00999)	(0.00988)	(0.00269)	(0.0118)
did (in SD)	-0.225***	-0.103*	-0.070	0.312***	-0.083*	-0.154***
	(0.050)	(0.055)	(0.049)	(0.070)	(0.049)	(0.058)
95% CI	(-0.046,-0.018)	(-0.012,0)	(-0.034,0.005)	(0.025,0.064)	(-0.01,0.001)	(-0.054,-0.008)
95% CI (in SD)	(-0.323,-0.127)	(-0.211,0.005)	(-0.166,0.026)	(0.175,0.449)	(-0.179,0.013)	(-0.268,-0.04)
Observations	8,174	8,174	8,174	8,174	8,174	8,174
R-squared	0.818	0.543	0.817	0.884	0.669	0.869
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1424	.0292	.3129	.1424	.0292	.3129
Pre-2007 SD	.1418	.0550	.2034	.1418	.0550	.2034
Schools	1296	1296	1296	1296	1296	1296

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 14: 4th Grade Math Test Scores, Black

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0746***	0.0473***	0.0977***	0.0359***	0.0382***	0.0326***
	(0.00645)	(0.00581)	(0.00856)	(0.00914)	(0.00832)	(0.0119)
did (in SD)	-0.604***	0.364***	0.455***	0.290***	0.294***	0.152***
	(0.052)	(0.045)	(0.040)	(0.074)	(0.064)	(0.055)
95% CI	(-0.087,-0.062)	(0.036,0.059)	(0.081,0.114)	(0.018,0.054)	(0.022,0.055)	(0.009,0.056)
95% CI (in SD)	(-0.706,-0.502)	(0.276,0.452)	(0.377,0.533)	(0.145,0.435)	(0.169,0.419)	(0.044,0.26)
Observations	11,703	11,703	11,703	11,703	11,703	11,703
R-squared	0.765	0.609	0.799	0.851	0.691	0.852
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1099	.1281	.6008	.1099	.1281	.6008
Pre-2007 SD	.1236	.1300	.2148	.1236	.1300	.2148
Schools	1686	1686	1686	1686	1686	1686

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 15: 8th Grade Math Test Scores, Black

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0773***	0.0125***	0.0457***	0.00950	0.00975	0.0109
	(0.0106)	(0.00334)	(0.0117)	(0.0149)	(0.00648)	(0.0150)
did (in SD)	-0.370***	0.216***	0.204***	0.045	0.169	0.049
	(0.051)	(0.058)	(0.052)	(0.071)	(0.112)	(0.067)
95% CI	(-0.098,-0.057)	(0.006,0.019)	(0.023,0.069)	(-0.02,0.039)	(-0.003,0.022)	(-0.019,0.04)
95% CI (in SD)	(-0.47,-0.27)	(0.102,0.33)	(0.102,0.306)	(-0.094,0.184)	(-0.051,0.389)	(-0.082,0.18)
Observations	8,087	8,087	8,087	8,087	8,087	8,087
R-squared	0.826	0.641	0.829	0.886	0.707	0.877
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.2902	.0303	.3162	.2902	.0303	.3162
Pre-2007 SD	.2089	.0578	.2236	.2089	.0578	.2236
Schools	1286	1286	1286	1286	1286	1286

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 16: 4th Grade ELA Test Scores, White

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0363***	0.0192***	0.0610***	0.0267***	-0.0233***	-0.0410***
	(0.00333)	(0.00606)	(0.00595)	(0.00462)	(0.00811)	(0.00806)
did (in SD)	-0.568***	0.129***	0.387***	0.418***	-0.157***	-0.260***
	(0.052)	(0.041)	(0.038)	(0.072)	(0.055)	(0.051)
95% CI	(-0.043,-0.03)	(0.007,0.031)	(0.049,0.073)	(0.018,0.036)	(-0.039,-0.007)	(-0.057,-0.025)
95% CI (in SD)	(-0.67,-0.466)	(0.049,0.209)	(0.313,0.461)	(0.277,0.559)	(-0.265,-0.049)	(-0.36,-0.16)
Observations	18,171	18,171	18,171	18,171	18,171	18,171
R-squared	0.725	0.739	0.859	0.828	0.821	0.894
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0443	.2029	.7424	.0443	.2029	.7424
Pre-2007 SD	.0639	.1484	.1575	.0639	.1484	.1575
Schools	2098	2098	2098	2098	2098	2098

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 17: 8th Grade ELA Test Scores, White

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0255***	-0.00226	0.0171**	0.0312***	-0.0307***	-0.0458***
	(0.00396)	(0.00630)	(0.00800)	(0.00604)	(0.00810)	(0.0106)
did (in SD)	-0.368***	-0.024	0.093**	0.450***	-0.328***	-0.249***
	(0.057)	(0.067)	(0.044)	(0.087)	(0.087)	(0.058)
95% CI	(-0.033,-0.018)	(-0.015,0.01)	(0.001,0.033)	(0.019,0.043)	(-0.047,-0.015)	(-0.067,-0.025)
95% CI (in SD)	(-0.48,-0.256)	(-0.155,0.107)	(0.007,0.179)	(0.279,0.621)	(-0.499,-0.157)	(-0.363,-0.135)
Observations	9,874	9,874	9,874	9,874	9,874	9,874
R-squared	0.765	0.729	0.862	0.847	0.785	0.892
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0503	.1105	.5757	.0503	.1105	.5757
Pre-2007 SD	.0693	.0935	.1837	.0693	.0935	.1837
Schools	1225	1225	1225	1225	1225	1225

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 18: 4th Grade Math Test Scores, White

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0382***	0.0918***	0.0930***	0.0135***	0.0558***	0.000619
	(0.00286)	(0.00690)	(0.00535)	(0.00409)	(0.0102)	(0.00723)
did (in SD)	-0.670***	0.508***	0.695***	0.237***	0.309***	0.005
	(0.050)	(0.038)	(0.040)	(0.072)	(0.056)	(0.054)
95% CI	(-0.044,-0.033)	(0.078,0.105)	(0.083,0.103)	(0.005,0.022)	(0.036,0.076)	(-0.014,0.015)
95% CI (in SD)	(-0.768,-0.572)	(0.434,0.582)	(0.617,0.773)	(0.096,0.378)	(0.199,0.419)	(-0.101,0.111)
Observations	18,235	18,235	18,235	18,235	18,235	18,235
R-squared	0.677	0.743	0.836	0.798	0.808	0.887
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0299	.3494	.8488	.0299	.3494	.8488
Pre-2007 SD	.0570	.1807	.1338	.0570	.1807	.1338
Schools	2112	2112	2112	2112	2112	2112

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 19: 8th Grade Math Test Scores, White

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0619***	0.0610***	0.115***	0.00984	0.0108	-0.0121
	(0.00462)	(0.0102)	(0.00998)	(0.00997)	(0.0131)	(0.0148)
did (in SD)	-0.582***	0.564***	0.567***	0.092	0.100	-0.060
	(0.043)	(0.094)	(0.049)	(0.094)	(0.121)	(0.073)
95% CI	(-0.071,-0.053)	(0.041,0.081)	(0.095,0.135)	(-0.01,0.029)	(-0.015,0.036)	(-0.041,0.017)
95% CI (in SD)	(-0.666,-0.498)	(0.38,0.748)	(0.471,0.663)	(-0.092,0.276)	(-0.137,0.337)	(-0.203,0.083)
Observations	9,721	10,104	9,721	9,721	10,104	9,721
R-squared	0.735	0.786	0.873	0.808	0.827	0.905
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0939	.1223	.6228	.0939	.1223	.6228
Pre-2007 SD	.1064	.1081	.2028	.1064	.1081	.2028
Schools	1226	1372	1226	1226	1372	1226

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 20: 4th Grade ELA Test Scores, Hispanic

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0657***	0.0440***	0.125***	0.0205***	0.0132***	0.0111
	(0.00548)	(0.00353)	(0.00736)	(0.00735)	(0.00395)	(0.0119)
did (in SD)	-0.567***	0.427***	0.603***	0.177***	0.128***	0.054
	(0.047)	(0.034)	(0.036)	(0.063)	(0.038)	(0.057)
95% CI	(-0.076,-0.055)	(0.037,0.051)	(0.111,0.139)	(0.006,0.035)	(0.005,0.021)	(-0.012,0.034)
95% CI (in SD)	(-0.659,-0.475)	(0.36,0.494)	(0.532,0.674)	(0.054,0.3)	(0.054,0.202)	(-0.058,0.166)
Observations	13,131	13,131	13,131	13,131	13,131	13,131
R-squared	0.792	0.502	0.791	0.859	0.627	0.841
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1119	.0818	.5244	.1119	.0818	.5244
Pre-2007 SD	.1159	.1030	.2073	.1159	.1030	.2073
Schools	1829	1829	1829	1829	1829	1829

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 21: 8th Grade ELA Test Scores, Hispanic

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0334***	0.00417*	0.0212**	0.00274	-0.00245	-0.0343**
	(0.00737)	(0.00236)	(0.00991)	(0.00841)	(0.00302)	(0.0134)
did (in SD)	-0.268***	0.072*	0.105**	0.022	-0.042	-0.171**
	(0.059)	(0.041)	(0.049)	(0.068)	(0.052)	(0.067)
95% CI	(-0.048,-0.019)	(0,0.009)	(0.002,0.041)	(-0.014,0.019)	(-0.008,0.003)	(-0.061,-0.008)
95% CI (in SD)	(-0.384,-0.152)	(-0.008,0.152)	(0.009,0.201)	(-0.111,0.155)	(-0.144,0.06)	(-0.302,-0.04)
Observations	8,378	8,378	8,378	8,378	8,378	8,378
R-squared	0.818	0.523	0.808	0.880	0.629	0.859
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1209	.0353	.3551	.1209	.0353	.3551
Pre-2007 SD	.1244	.0579	.2010	.1244	.0579	.2010
Schools	1266	1266	1266	1266	1266	1266

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 22: 4th Grade Math Test Scores, Hispanic

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0614***	0.0725***	0.117***	0.0136*	0.0505***	0.0460***
	(0.00550)	(0.00645)	(0.00825)	(0.00826)	(0.00969)	(0.0106)
did (in SD)	-0.571***	0.494***	0.570***	0.127*	0.344***	0.224***
	(0.051)	(0.044)	(0.040)	(0.077)	(0.066)	(0.052)
95% CI	(-0.072,-0.051)	(0.06,0.085)	(0.101,0.133)	(-0.003,0.03)	(0.032,0.069)	(0.025,0.067)
95% CI (in SD)	(-0.671,-0.471)	(0.408,0.58)	(0.492,0.648)	(-0.024,0.278)	(0.215,0.473)	(0.122,0.326)
Observations	13,379	13,379	13,379	13,379	13,379	13,379
R-squared	0.775	0.606	0.821	0.853	0.686	0.869
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0944	.1658	.6617	.0944	.1658	.6617
Pre-2007 SD	.1075	.1467	.2051	.1075	.1467	.2051
Schools	1833	1833	1833	1833	1833	1833

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 23: 8th Grade Math Test Scores, Hispanic

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0859***	0.0155***	0.0873***	-0.0165	-0.00829	0.0134
	(0.0101)	(0.00452)	(0.0136)	(0.0137)	(0.00775)	(0.0171)
did (in SD)	-0.455***	0.229***	0.385***	-0.087	-0.122	0.059
	(0.054)	(0.067)	(0.060)	(0.073)	(0.114)	(0.075)
95% CI	(-0.106,-0.066)	(0.007,0.024)	(0.061,0.114)	(-0.043,0.01)	(-0.023,0.007)	(-0.02,0.047)
95% CI (in SD)	(-0.561,-0.349)	(0.098,0.36)	(0.267,0.503)	(-0.23,0.056)	(-0.345,0.101)	(-0.088,0.206)
Observations	8,282	8,282	8,282	8,282	8,282	8,282
R-squared	0.812	0.621	0.835	0.875	0.695	0.882
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.2469	.0419	.3702	.2469	.0419	.3702
Pre-2007 SD	.1886	.0677	.2268	.1886	.0677	.2268
Schools	1256	1256	1256	1256	1256	1256

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 24: 4th Grade ELA Test Scores, Asian

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	0.00225	0.0626***	0.0377***	-0.00331	0.000977	0.00336
	(0.00653)	(0.0101)	(0.00966)	(0.00771)	(0.0145)	(0.0141)
did (in SD)	0.037	0.321***	0.225***	-0.055	0.005	0.020
	(0.108)	(0.052)	(0.058)	(0.128)	(0.074)	(0.084)
95% CI	(-0.011,0.015)	(0.043,0.082)	(0.019,0.057)	(-0.018,0.012)	(-0.027,0.029)	(-0.024,0.031)
95% CI (in SD)	(-0.175,0.249)	(0.219,0.423)	(0.111,0.339)	(-0.306,0.196)	(-0.14,0.15)	(-0.145,0.185)
Observations	6,189	6,189	6,189	6,189	6,189	6,189
R-squared	0.693	0.659	0.780	0.802	0.738	0.836
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0303	.2594	.7977	.0303	.2594	.7977
Pre-2007 SD	.0604	.1951	.1675	.0604	.1951	.1675
Schools	1023	1023	1023	1023	1023	1023

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 25: 8th Grade ELA Test Scores, Asian

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	0.0163	0.00790	-0.00187	-0.00857	0.0234*	-0.0211
	(0.0103)	(0.00913)	(0.0119)	(0.00906)	(0.0135)	(0.0162)
did (in SD)	0.219	0.051	-0.009	-0.115	0.152*	-0.097
	(0.138)	(0.059)	(0.055)	(0.122)	(0.088)	(0.075)
95% CI	(-0.004,0.036)	(-0.01,0.026)	(-0.025,0.021)	(-0.026,0.009)	(-0.003,0.05)	(-0.053,0.011)
95% CI (in SD)	(-0.051,0.489)	(-0.065,0.167)	(-0.117,0.099)	(-0.354,0.124)	(-0.02,0.324)	(-0.244,0.05)
Observations	4,215	4,215	4,215	4,215	4,215	4,215
R-squared	0.735	0.712	0.834	0.850	0.785	0.880
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0398	.1676	.6643	.0398	.1676	.6643
Pre-2007 SD	.0744	.1538	.2167	.0744	.1538	.2167
Schools	656	656	656	656	656	656

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 26: 4th Grade Math Test Scores, Asian

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0166***	0.0909***	0.0474***	-0.00760	0.0189	0.0242**
	(0.00526)	(0.0111)	(0.00926)	(0.00686)	(0.0181)	(0.0114)
did (in SD)	-0.307***	0.407***	0.363***	-0.140	0.085	0.186**
	(0.097)	(0.050)	(0.071)	(0.127)	(0.081)	(0.087)
95% CI	(-0.027,-0.006)	(0.069,0.113)	(0.029,0.066)	(-0.021,0.006)	(-0.017,0.054)	(0.002,0.047)
95% CI (in SD)	(-0.497,-0.117)	(0.309,0.505)	(0.224,0.502)	(-0.389,0.109)	(-0.074,0.244)	(0.015,0.357)
Observations	6,395	6,395	6,395	6,395	6,395	6,395
R-squared	0.647	0.731	0.755	0.772	0.794	0.827
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0236	.4891	.8943	.0236	.4891	.8943
Pre-2007 SD	.0541	.2236	.1304	.0541	.2236	.1304
Schools	1041	1041	1041	1041	1041	1041

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 27: 8th Grade Math Test Scores, Asian

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0485***	0.0871***	0.109***	-0.0202*	-0.0240	-0.00116
	(0.00595)	(0.0136)	(0.0109)	(0.0114)	(0.0230)	(0.0218)
did (in SD)	-0.507***	0.437***	0.529***	-0.211*	-0.120	-0.006
	(0.062)	(0.068)	(0.053)	(0.119)	(0.115)	(0.106)
95% CI	(-0.06,-0.037)	(0.06,0.114)	(0.088,0.13)	(-0.043,0.002)	(-0.069,0.021)	(-0.044,0.042)
95% CI (in SD)	(-0.629,-0.385)	(0.304,0.57)	(0.425,0.633)	(-0.444,0.022)	(-0.345,0.105)	(-0.214,0.202)
Observations	4,084	4,084	4,084	4,084	4,084	4,084
R-squared	0.701	0.795	0.807	0.790	0.849	0.861
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0629	.2930	.7585	.0629	.2930	.7585
Pre-2007 SD	.0957	.1994	.2059	.0957	.1994	.2059
Schools	656	656	656	656	656	656

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 28: 4th Grade ELA Test Scores, Native American

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.414***	0.0803**	0.0480	-0.246**	-0.347***	-0.142
	(0.0483)	(0.0325)	(0.0930)	(0.116)	(0.0623)	(0.117)
did (in SD)	-3.160***	0.839**	0.216	-1.878**	-3.626***	-0.638
	(0.369)	(0.340)	(0.418)	(0.885)	(0.651)	(0.526)
95% CI	(-0.509,-0.319)	(0.017,0.144)	(-0.134,0.23)	(-0.473,-0.019)	(-0.469,-0.225)	(-0.371,0.087)
95% CI (in SD)	(-3.883,-2.437)	(0.173,1.505)	(-0.603,1.035)	(-3.613,-0.143)	(-4.902,-2.35)	(-1.669,0.393)
Observations	172	172	172	172	172	172
R-squared	0.800	0.760	0.788	0.864	0.895	0.841
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0988	.0717	.5083	.0988	.0717	.5083
Pre-2007 SD	.1310	.0957	.2224	.1310	.0957	.2224
Schools	69	69	69	69	69	69

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 29: 8th Grade ELA Test Scores, Native American

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.213***	0.0275	0.238***	0.230	0.0667	0.263
	(0.0694)	(0.0180)	(0.0582)	(0.373)	(0.127)	(0.253)
did (in SD)	-1.378***	0.455	1.200***	1.488	1.104	1.326
	(0.449)	(0.298)	(0.293)	(2.413)	(2.103)	(1.275)
95% CI	(-0.349,-0.077)	(-0.008,0.063)	(0.124,0.352)	(-0.501,0.961)	(-0.182,0.316)	(-0.233,0.759)
95% CI (in SD)	(-2.258,-0.498)	(-0.129,1.039)	(0.626,1.774)	(-3.241,6.217)	(-3.018,5.226)	(-1.173,3.825)
Observations	260	260	260	260	260	260
R-squared	0.759	0.355	0.684	0.823	0.521	0.736
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1144	.0268	.2991	.1144	.0268	.2991
Pre-2007 SD	.1546	.0604	.1984	.1546	.0604	.1984
Schools	83	83	83	83	83	83

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 30: 4th Grade Math Test Scores, Native American

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.272***	-0.301***	-0.272***	-0.333***	-0.552***	0.0177
	(0.0603)	(0.0689)	(0.0935)	(0.0981)	(0.0409)	(0.115)
did (in SD)	-2.598***	-2.205***	-1.166***	-3.181***	-4.044***	0.076
	(0.576)	(0.505)	(0.401)	(0.937)	(0.300)	(0.493)
95% CI	(-0.39,-0.154)	(-0.436,-0.166)	(-0.455,-0.089)	(-0.525,-0.141)	(-0.632,-0.472)	(-0.208,0.243)
95% CI (in SD)	(-3.727,-1.469)	(-3.195,-1.215)	(-1.952,-0.38)	(-5.018,-1.344)	(-4.632,-3.456)	(-0.89,1.042)
Observations	163	163	163	163	163	163
R-squared	0.820	0.758	0.846	0.874	0.852	0.905
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0646	.1479	.7176	.0646	.1479	.7176
Pre-2007 SD	.1047	.1365	.2333	.1047	.1365	.2333
Schools	65	65	65	65	65	65

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 31: 8th Grade Math Test Scores, Native American

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.264***	-0.0516	0.230***	-0.0816	-0.120	0.00718
	(0.0538)	(0.0392)	(0.0775)	(0.156)	(0.0975)	(0.130)
did (in SD)	-1.473***	-0.622	1.031***	-0.455	-1.446	0.032
	(0.300)	(0.472)	(0.347)	(0.871)	(1.175)	(0.583)
95% CI	(-0.369,-0.159)	(-0.128,0.025)	(0.078,0.382)	(-0.387,0.224)	(-0.311,0.071)	(-0.248,0.262)
95% CI (in SD)	(-2.061,-0.885)	(-1.547,0.303)	(0.351,1.711)	(-2.162,1.252)	(-3.749,0.857)	(-1.111,1.175)
Observations	250	250	250	250	250	250
R-squared	0.804	0.598	0.817	0.848	0.659	0.859
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1817	.0425	.3978	.1817	.0425	.3978
Pre-2007 SD	.1792	.0830	.2231	.1792	.0830	.2231
Schools	81	81	81	81	81	81

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 32: 4th Grade ELA Test Scores, Economically Advantaged

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0410***	0.0303***	0.105***	-0.0295***	-0.0185**	0.0365***
	(0.00281)	(0.00537)	(0.00543)	(0.00627)	(0.00767)	(0.00989)
did (in SD)	-0.379***	0.205***	0.528***	-0.272***	-0.125**	0.184***
	(0.026)	(0.036)	(0.027)	(0.058)	(0.052)	(0.050)
95% CI	(-0.047,-0.035)	(0.02,0.041)	(0.094,0.116)	(-0.042,-0.017)	(-0.034,-0.003)	(0.017,0.056)
95% CI (in SD)	(-0.43,-0.328)	(0.134,0.276)	(0.475,0.581)	(-0.386,-0.158)	(-0.227,-0.023)	(0.086,0.282)
Observations	25,378	25,378	25,378	25,378	25,378	25,378
R-squared	0.664	0.721	0.804	0.752	0.798	0.846
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0584	.1878	.7133	.0584	.1878	.7133
Pre-2007 SD	.1083	.1476	.1988	.1083	.1476	.1988
Schools	2694	2694	2694	2694	2694	2694

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 33: 4th Grade ELA Test Scores, Economically Disadvantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0581***	0.00874***	0.0658***	0.0760***	-0.0126***	-0.0352***
	(0.00333)	(0.00212)	(0.00462)	(0.00437)	(0.00258)	(0.00645)
did (in SD)	-0.619***	0.093***	0.360***	0.809***	-0.135***	-0.193***
	(0.035)	(0.023)	(0.025)	(0.047)	(0.028)	(0.035)
95% CI	(-0.065,-0.052)	(0.005,0.013)	(0.057,0.075)	(0.067,0.085)	(-0.018,-0.008)	(-0.048,-0.023)
95% CI (in SD)	(-0.688,-0.55)	(0.048,0.138)	(0.311,0.409)	(0.717,0.901)	(-0.19,-0.08)	(-0.262,-0.124)
Observations	26,142	26,142	26,142	26,142	26,142	26,142
R-squared	0.784	0.564	0.799	0.854	0.661	0.843
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0948	.0813	.5510	.0948	.0813	.5510
Pre-2007 SD	.0939	.0935	.1827	.0939	.0935	.1827
Schools	2631	2631	2631	2631	2631	2631

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 34: 8th Grade ELA Test Scores, Economically Advantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0382***	0.000682	0.0375***	-0.0404***	-0.0183**	0.00531
	(0.00392)	(0.00526)	(0.00760)	(0.00895)	(0.00735)	(0.0108)
did (in SD)	-0.332***	0.008	0.178***	-0.351***	-0.211**	0.025
	(0.034)	(0.061)	(0.036)	(0.078)	(0.085)	(0.051)
95% CI	(-0.046,-0.031)	(-0.01,0.011)	(0.023,0.052)	(-0.058,-0.023)	(-0.033,-0.004)	(-0.016,0.026)
95% CI (in SD)	(-0.399,-0.265)	(-0.112,0.128)	(0.107,0.249)	(-0.504,-0.198)	(-0.378,-0.044)	(-0.075,0.125)
Observations	14,875	14,875	14,875	14,875	14,875	14,875
R-squared	0.714	0.718	0.834	0.789	0.776	0.870
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0686	.0963	.5362	.0686	.0963	.5362
Pre-2007 SD	.1151	.0869	.2108	.1151	.0869	.2108
Schools	1755	1755	1755	1755	1755	1755

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 35: 8th Grade ELA Test Scores, Economically Disadvantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0268***	-0.00275	-0.00768	0.0758***	-0.0164***	-0.0508***
	(0.00501)	(0.00225)	(0.00685)	(0.00636)	(0.00284)	(0.00814)
did (in SD)	-0.256***	-0.051	-0.043	0.724***	-0.304***	-0.286***
	(0.048)	(0.042)	(0.039)	(0.061)	(0.053)	(0.046)
95% CI	(-0.037,-0.017)	(-0.007,0.002)	(-0.021,0.006)	(0.063,0.088)	(-0.022,-0.011)	(-0.067,-0.035)
95% CI (in SD)	(-0.35,-0.162)	(-0.133,0.031)	(-0.119,0.033)	(0.604,0.844)	(-0.408,-0.2)	(-0.376,-0.196)
Observations	15,333	15,333	15,333	15,333	15,333	15,333
R-squared	0.812	0.620	0.828	0.882	0.715	0.872
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1127	.0337	.3503	.1127	.0337	.3503
Pre-2007 SD	.1047	.0539	.1778	.1047	.0539	.1778
Schools	1748	1748	1748	1748	1748	1748

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 36: 4th Grade Math Test Scores, Economically Advantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0344***	0.110***	0.110***	-0.0361***	0.0812***	0.0672***
	(0.00286)	(0.00619)	(0.00491)	(0.00608)	(0.00918)	(0.00934)
did (in SD)	-0.340***	0.590***	0.596***	-0.357***	0.436***	0.364***
	(0.028)	(0.033)	(0.027)	(0.060)	(0.049)	(0.051)
95% CI	(-0.04,-0.029)	(0.098,0.122)	(0.1,0.12)	(-0.048,-0.024)	(0.063,0.099)	(0.049,0.086)
95% CI (in SD)	(-0.395,-0.285)	(0.525,0.655)	(0.543,0.649)	(-0.475,-0.239)	(0.34,0.532)	(0.264,0.464)
Observations	25,469	25,469	25,469	25,469	25,469	25,469
R-squared	0.645	0.706	0.791	0.751	0.775	0.846
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0471	.3185	.8132	.0471	.3185	.8132
Pre-2007 SD	.1012	.1864	.1847	.1012	.1864	.1847
Schools	2704	2704	2704	2704	2704	2704

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 37: 4th Grade Math Test Scores, Economically Disadvantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0667***	0.0621***	0.112***	0.0512***	0.0567***	0.00911
	(0.00347)	(0.00375)	(0.00513)	(0.00509)	(0.00568)	(0.00655)
did (in SD)	-0.797***	0.431***	0.618***	0.612***	0.394***	0.050
	(0.041)	(0.026)	(0.028)	(0.061)	(0.039)	(0.036)
95% CI	(-0.074,-0.06)	(0.055,0.069)	(0.102,0.122)	(0.041,0.061)	(0.046,0.068)	(-0.004,0.022)
95% CI (in SD)	(-0.877,-0.717)	(0.38,0.482)	(0.563,0.673)	(0.492,0.732)	(0.318,0.47)	(-0.021,0.121)
Observations	26,256	26,256	26,256	26,256	26,256	26,256
R-squared	0.762	0.706	0.825	0.845	0.762	0.869
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0690	.1774	.7015	.0690	.1774	.7015
Pre-2007 SD	.0837	.1440	.1811	.0837	.1440	.1811
Schools	2639	2639	2639	2639	2639	2639

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 38: 8th Grade Math Test Scores, Economically Advantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0869***	0.0590***	0.141***	-0.0767***	-0.00823	0.0346***
	(0.00614)	(0.00847)	(0.00901)	(0.0130)	(0.0118)	(0.0126)
did (in SD)	-0.509***	0.567***	0.583***	-0.450***	-0.079	0.143***
	(0.036)	(0.081)	(0.037)	(0.076)	(0.113)	(0.052)
95% CI	(-0.099,-0.075)	(0.042,0.076)	(0.123,0.159)	(-0.102,-0.051)	(-0.031,0.015)	(0.01,0.059)
95% CI (in SD)	(-0.58,-0.438)	(0.408,0.726)	(0.51,0.656)	(-0.599,-0.301)	(-0.3,0.142)	(0.041,0.245)
Observations	14,711	14,711	14,711	14,711	14,711	14,711
R-squared	0.746	0.779	0.860	0.807	0.824	0.894
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1309	.1102	.5709	.1309	.1102	.5709
Pre-2007 SD	.1706	.1040	.2420	.1706	.1040	.2420
Schools	1756	1756	1756	1756	1756	1756

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 39: 8th Grade Math Test Scores, Economically Disadvantaged

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0809***	0.0420***	0.0821***	0.0344***	0.0339***	-0.0189*
	(0.00622)	(0.00527)	(0.00810)	(0.0103)	(0.00711)	(0.0110)
did (in SD)	-0.482***	0.594***	0.386***	0.205***	0.479***	-0.089*
	(0.037)	(0.075)	(0.038)	(0.061)	(0.101)	(0.052)
95% CI	(-0.093,-0.069)	(0.032,0.052)	(0.066,0.098)	(0.014,0.055)	(0.02,0.048)	(-0.04,0.003)
95% CI (in SD)	(-0.555,-0.409)	(0.447,0.741)	(0.312,0.46)	(0.085,0.325)	(0.281,0.677)	(-0.191,0.013)
Observations	15,209	15,209	15,209	15,209	15,209	15,209
R-squared	0.815	0.788	0.852	0.876	0.842	0.894
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.2040	.0464	.4077	.2040	.0464	.4077
Pre-2007 SD	.1677	.0707	.2126	.1677	.0707	.2126
Schools	1745	1745	1745	1745	1745	1745

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 40: 4th Grade ELA Test Scores, Male

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(1) Level 1	(2) Level 4	(3) Level 34
did	-0.0387***	0.0547***	0.0987***	0.00110	0.00693**	-0.0241***
	(0.00308)	(0.00268)	(0.00378)	(0.00427)	(0.00286)	(0.00505)
did (in SD)	-0.394***	0.444***	0.473***	0.011	0.056**	-0.115***
	(0.031)	(0.022)	(0.018)	(0.043)	(0.023)	(0.024)
95% CI	(-0.045,-0.033)	(0.049,0.06)	(0.091,0.106)	(-0.007,0.009)	(0.001,0.013)	(-0.034,-0.014)
95% CI (in SD)	(-0.455,-0.333)	(0.401,0.487)	(0.438,0.508)	(-0.073,0.095)	(0.011,0.101)	(-0.162,-0.068)
Observations	32,145	32,145	32,145	32,145	32,145	32,145
R-squared	0.760	0.673	0.854	0.836	0.762	0.887
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0851	.1253	.6218	.0851	.1253	.6218
Pre-2007 SD	.0983	.1231	.2087	.0983	.1231	.2087
Schools	2742	2742	2742	2742	2742	2742

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 41: 4th Grade ELA Test Scores, Female

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0221***	0.0693***	0.0832***	0.00288	0.00569*	-0.0128**
	(0.00260)	(0.00289)	(0.00373)	(0.00373)	(0.00343)	(0.00511)
did (in SD)	-0.289***	0.462***	0.419***	0.038	0.038*	-0.064**
	(0.034)	(0.019)	(0.019)	(0.049)	(0.023)	(0.026)
95% CI	(-0.027,-0.017)	(0.064,0.075)	(0.076,0.091)	(-0.004,0.01)	(-0.001,0.012)	(-0.023,-0.003)
95% CI (in SD)	(-0.356,-0.222)	(0.425,0.499)	(0.382,0.456)	(-0.058,0.134)	(-0.007,0.083)	(-0.115,-0.013)
Observations	32,134	32,134	32,134	32,134	32,134	32,134
R-squared	0.715	0.725	0.850	0.807	0.792	0.883
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0580	.1809	.6807	.0580	.1809	.6807
Pre-2007 SD	.0766	.1500	.1985	.0766	.1500	.1985
Schools	2736	2736	2736	2736	2736	2736

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 42: 8th Grade ELA Test Scores, Male

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0262***	0.0115***	0.0321***	-0.00698	-0.000715	-0.0367***
	(0.00421)	(0.00236)	(0.00584)	(0.00635)	(0.00352)	(0.00679)
did (in SD)	-0.227***	0.168***	0.152***	-0.060	-0.010	-0.174***
	(0.036)	(0.035)	(0.028)	(0.055)	(0.051)	(0.032)
95% CI	(-0.034,-0.018)	(0.007,0.016)	(0.021,0.044)	(-0.019,0.005)	(-0.008,0.006)	(-0.05,-0.023)
95% CI (in SD)	(-0.298,-0.156)	(0.099,0.237)	(0.097,0.207)	(-0.168,0.048)	(-0.11,0.09)	(-0.237,-0.111)
Observations	17,233	17,233	17,233	17,233	17,233	17,233
R-squared	0.805	0.683	0.885	0.868	0.739	0.911
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1009	.0585	.4259	.1009	.0585	.4259
Pre-2007 SD	.1155	.0684	.2114	.1155	.0684	.2114
Schools	1807	1807	1807	1807	1807	1807

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 43: 8th Grade ELA Test Scores, Female

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.00484	0.0137***	0.0129**	-0.00956*	0.00162	-0.0281***
	(0.00348)	(0.00338)	(0.00597)	(0.00508)	(0.00381)	(0.00635)
did (in SD)	-0.056	0.139***	0.060**	-0.110*	0.016	-0.130***
	(0.040)	(0.034)	(0.028)	(0.058)	(0.039)	(0.029)
95% CI	(-0.012,0.002)	(0.007,0.02)	(0.001,0.025)	(-0.02,0)	(-0.006,0.009)	(-0.041,-0.016)
95% CI (in SD)	(-0.134,0.022)	(0.072,0.206)	(0.005,0.115)	(-0.224,0.004)	(-0.06,0.092)	(-0.187,-0.073)
Observations	17,212	17,212	17,212	17,212	17,212	17,212
R-squared	0.745	0.730	0.890	0.830	0.785	0.916
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0622	.0996	.5310	.0622	.0996	.5310
Pre-2007 SD	.0870	.0989	.2157	.0870	.0989	.2157
Schools	1804	1804	1804	1804	1804	1804

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 44: 4th Grade Math Test Scores, Male

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0402***	0.0982***	0.114***	-0.0306***	0.0285***	0.0424***
	(0.00296)	(0.00363)	(0.00423)	(0.00440)	(0.00509)	(0.00516)
did (in SD)	-0.463***	0.522***	0.600***	-0.352***	0.151***	0.223***
	(0.034)	(0.019)	(0.022)	(0.051)	(0.027)	(0.027)
95% CI	(-0.046,-0.034)	(0.091,0.105)	(0.106,0.122)	(-0.039,-0.022)	(0.019,0.038)	(0.032,0.053)
95% CI (in SD)	(-0.53,-0.396)	(0.485,0.559)	(0.557,0.643)	(-0.452,-0.252)	(0.098,0.204)	(0.17,0.276)
Observations	32,147	32,147	32,147	32,147	32,147	32,147
R-squared	0.708	0.764	0.835	0.806	0.814	0.878
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0590	.2856	.7714	.0590	.2856	.7714
Pre-2007 SD	.0869	.1883	.1900	.0869	.1883	.1900
Schools	2745	2745	2745	2745	2745	2745

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 45: 4th Grade Math Test Scores, Female

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0289***	0.0932***	0.107***	-0.0210***	0.0324***	0.0388***
	(0.00290)	(0.00373)	(0.00427)	(0.00441)	(0.00534)	(0.00535)
did (in SD)	-0.366***	0.517***	0.563***	-0.266***	0.180***	0.204***
	(0.037)	(0.021)	(0.022)	(0.056)	(0.030)	(0.028)
95% CI	(-0.035,-0.023)	(0.086,0.101)	(0.099,0.115)	(-0.03,-0.012)	(0.022,0.043)	(0.028,0.049)
95% CI (in SD)	(-0.439,-0.293)	(0.476,0.558)	(0.52,0.606)	(-0.376,-0.156)	(0.121,0.239)	(0.149,0.259)
Observations	32,132	32,132	32,132	32,132	32,132	32,132
R-squared	0.695	0.742	0.833	0.798	0.795	0.875
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0528	.2615	.7627	.0528	.2615	.7627
Pre-2007 SD	.0790	.1804	.1899	.0790	.1804	.1899
Schools	2736	2736	2736	2736	2736	2736

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. The regression presented here controls for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 46: 8th Grade Math Test Scores, Male

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.111***	0.0405***	0.142***	-0.0554***	-0.0240***	0.00883
	(0.00537)	(0.00511)	(0.00685)	(0.00925)	(0.00709)	(0.00866)
did (in SD)	-0.643***	0.396***	0.583***	-0.321***	-0.234***	0.036
	(0.031)	(0.050)	(0.028)	(0.054)	(0.069)	(0.036)
95% CI	(-0.122,-0.1)	(0.03,0.051)	(0.129,0.155)	(-0.074,-0.037)	(-0.038,-0.01)	(-0.008,0.026)
95% CI (in SD)	(-0.704,-0.582)	(0.298,0.494)	(0.528,0.638)	(-0.427,-0.215)	(-0.369,-0.099)	(-0.035,0.107)
Observations	17,098	17,098	17,098	17,098	17,098	17,098
R-squared	0.811	0.788	0.884	0.863	0.832	0.916
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1651	.0960	.5146	.1651	.0960	.5146
Pre-2007 SD	.1727	.1024	.2437	.1727	.1024	.2437
Schools	1816	1816	1816	1816	1816	1816

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 47: 8th Grade Math Test Scores, Female

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0925***	0.0356***	0.125***	-0.0549***	-0.0286***	0.0119
	(0.00495)	(0.00574)	(0.00710)	(0.00901)	(0.00741)	(0.00868)
did (in SD)	-0.549***	0.352***	0.510***	-0.326***	-0.283***	0.049
	(0.029)	(0.057)	(0.029)	(0.053)	(0.073)	(0.035)
95% CI	(-0.102,-0.083)	(0.024,0.047)	(0.111,0.139)	(-0.073,-0.037)	(-0.043,-0.014)	(-0.005,0.029)
95% CI (in SD)	(-0.606,-0.492)	(0.24,0.464)	(0.453,0.567)	(-0.43,-0.222)	(-0.426,-0.14)	(-0.02,0.118)
Observations	17,072	17,072	17,072	17,072	17,072	17,072
R-squared	0.792	0.779	0.881	0.845	0.827	0.912
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1502	.0943	.5255	.1502	.0943	.5255
Pre-2007 SD	.1685	.1011	.2450	.1685	.1011	.2450
Schools	1813	1813	1813	1813	1813	1813

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 48: 4th Grade ELA Test Scores, Disabled

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.190***	0.0199***	0.139***	0.0174*	-0.00134	0.0145*
	(0.00650)	(0.00130)	(0.00515)	(0.00907)	(0.00148)	(0.00833)
did (in SD)	-0.837***	0.298***	0.595***	0.077*	-0.020	0.062*
	(0.029)	(0.019)	(0.022)	(0.040)	(0.022)	(0.036)
95% CI	(-0.203,-0.177)	(0.017,0.022)	(0.129,0.149)	(0,0.035)	(-0.004,0.002)	(-0.002,0.031)
95% CI (in SD)	(-0.894,-0.78)	(0.261,0.335)	(0.552,0.638)	(-0.001,0.155)	(-0.063,0.023)	(-0.009,0.133)
Observations	26,617	26,617	26,617	26,617	26,617	26,617
R-squared	0.705	0.290	0.612	0.771	0.451	0.691
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.2753	.0241	.2857	.2753	.0241	.2857
Pre-2007 SD	.2269	.0668	.2336	.2269	.0668	.2336
Schools	2642	2642	2642	2642	2642	2642

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 49: 4th Grade ELA Test Scores, General Education

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0200***	0.0696***	0.0950***	0.00434	0.00514	-0.0147***
	(0.00257)	(0.00297)	(0.00378)	(0.00368)	(0.00328)	(0.00492)
did (in SD)	-0.341***	0.505***	0.484***	0.074	0.037	-0.075***
	(0.044)	(0.022)	(0.019)	(0.063)	(0.024)	(0.025)
95% CI	(-0.025,-0.015)	(0.064,0.075)	(0.088,0.102)	(-0.003,0.012)	(-0.001,0.012)	(-0.024,-0.005)
95% CI (in SD)	(-0.427,-0.255)	(0.462,0.548)	(0.447,0.521)	(-0.049,0.197)	(-0.01,0.084)	(-0.124,-0.026)
Observations	27,236	27,236	27,236	27,236	27,236	27,236
R-squared	0.735	0.762	0.888	0.837	0.835	0.919
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0406	.1665	.6984	.0406	.1665	.6984
Pre-2007 SD	.0587	.1377	.1962	.0587	.1377	.1962
Schools	2684	2684	2684	2684	2684	2684

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 50: 8th Grade ELA Test Scores, Disabled

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.164***	0.00297***	0.0387***	-0.0108	0.000972	-0.0187***
	(0.00761)	(0.000563)	(0.00384)	(0.01000)	(0.000904)	(0.00722)
did (in SD)	-0.763***	0.151***	0.300***	-0.050	0.049	-0.145***
	(0.035)	(0.029)	(0.030)	(0.047)	(0.046)	(0.056)
95% CI	(-0.179,-0.149)	(0.002,0.004)	(0.031,0.046)	(-0.03,0.009)	(-0.001,0.003)	(-0.033,-0.005)
95% CI (in SD)	(-0.832,-0.694)	(0.094,0.208)	(0.241,0.359)	(-0.142,0.042)	(-0.041,0.139)	(-0.255,-0.035)
Observations	14,957	14,957	14,957	14,957	14,957	14,957
R-squared	0.814	0.236	0.650	0.856	0.376	0.700
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.3056	.0046	.1108	.3056	.0046	.1108
Pre-2007 SD	.2150	.0197	.1291	.2150	.0197	.1291
Schools	1699	1699	1699	1699	1699	1699

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 51: 8th Grade ELA Test Scores, General Education

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.00146	0.0144***	0.0254***	-0.0132**	0.00122	-0.0328***
	(0.00375)	(0.00324)	(0.00625)	(0.00556)	(0.00412)	(0.00677)
did (in SD)	-0.020	0.166***	0.117***	-0.183**	0.014	-0.151***
	(0.052)	(0.037)	(0.029)	(0.077)	(0.047)	(0.031)
95% CI	(-0.009,0.006)	(0.008,0.021)	(0.013,0.038)	(-0.024,-0.002)	(-0.007,0.009)	(-0.046,-0.02)
95% CI (in SD)	(-0.122,0.082)	(0.093,0.239)	(0.06,0.174)	(-0.334,-0.032)	(-0.078,0.106)	(-0.212,-0.09)
Observations	15,451	15,451	15,451	15,451	15,451	15,451
R-squared	0.746	0.754	0.910	0.842	0.807	0.934
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0444	.0903	.5338	.0444	.0903	.5338
Pre-2007 SD	.0721	.0870	.2176	.0721	.0870	.2176
Schools	1780	1780	1780	1780	1780	1780

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 52: 4th Grade Math Test Scores, Disabled

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(1) Level 1	(2) Level 4	(3) Level 34
did	-0.209***	0.0697***	0.236***	-0.0382***	0.0413***	0.116***
	(0.00652)	(0.00310)	(0.00636)	(0.00872)	(0.00462)	(0.00980)
did (in SD)	-1.000***	0.551***	0.877***	-0.183***	0.326***	0.431***
	(0.031)	(0.024)	(0.024)	(0.042)	(0.036)	(0.036)
95% CI	(-0.222,-0.196)	(0.064,0.076)	(0.224,0.248)	(-0.055,-0.021)	(0.032,0.05)	(0.097,0.135)
95% CI (in SD)	(-1.061,-0.939)	(0.504,0.598)	(0.83,0.924)	(-0.265,-0.101)	(0.255,0.397)	(0.36,0.502)
Observations	26,659	26,659	26,659	26,659	26,659	26,659
R-squared	0.708	0.454	0.697	0.780	0.557	0.756
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.2000	.0807	.4851	.2000	.0807	.4851
Pre-2007 SD	.2089	.1266	.2691	.2089	.1266	.2691
Schools	2649	2649	2649	2649	2649	2649

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 53: 4th Grade Math Test Scores, General Education

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.0196***	0.105***	0.104***	-0.0154***	0.0308***	0.0286***
	(0.00270)	(0.00400)	(0.00437)	(0.00424)	(0.00545)	(0.00516)
did (in SD)	-0.338***	0.564***	0.584***	-0.266***	0.165***	0.160***
	(0.047)	(0.021)	(0.025)	(0.073)	(0.029)	(0.029)
95% CI	(-0.025,-0.014)	(0.097,0.113)	(0.095,0.113)	(-0.024,-0.007)	(0.02,0.041)	(0.018,0.039)
95% CI (in SD)	(-0.43,-0.246)	(0.523,0.605)	(0.535,0.633)	(-0.409,-0.123)	(0.108,0.222)	(0.103,0.217)
Observations	27,267	27,267	27,267	27,267	27,267	27,267
R-squared	0.681	0.804	0.854	0.801	0.854	0.898
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0342	.2963	.8038	.0342	.2963	.8038
Pre-2007 SD	.0580	.1862	.1782	.0580	.1862	.1782
Schools	2689	2689	2689	2689	2689	2689

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 54: 8th Grade Math Test Scores, Disabled

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Level 1	Level 4	Level 34	Level 1	Level 4	Level 34
did	-0.236***	0.00632***	0.103***	-0.123***	-0.000451	-0.00207
	(0.00901)	(0.00135)	(0.00714)	(0.0125)	(0.00259)	(0.0128)
did (in SD)	-0.923***	0.235***	0.545***	-0.481***	-0.017	-0.011
	(0.035)	(0.050)	(0.038)	(0.049)	(0.096)	(0.068)
95% CI	(-0.254,-0.218)	(0.004,0.009)	(0.089,0.117)	(-0.148,-0.099)	(-0.006,0.005)	(-0.027,0.023)
95% CI (in SD)	(-0.992,-0.854)	(0.137,0.333)	(0.471,0.619)	(-0.577,-0.385)	(-0.205,0.171)	(-0.144,0.122)
Observations	14,866	14,866	14,866	14,866	14,866	14,866
R-squared	0.796	0.372	0.728	0.844	0.434	0.780
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.4259	.0084	.1904	.4259	.0084	.1904
Pre-2007 SD	.2557	.0269	.1890	.2557	.0269	.1890
Schools	1696	1696	1696	1696	1696	1696

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 55: 8th Grade Math Test Scores, General Education

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0941***	0.0399***	0.145***	-0.0579***	-0.0322***	0.0230**
	(0.00515)	(0.00521)	(0.00742)	(0.00946)	(0.00771)	(0.00925)
did (in SD)	-0.615***	0.374***	0.581***	-0.378***	-0.302***	0.092**
	(0.034)	(0.049)	(0.030)	(0.062)	(0.072)	(0.037)
95% CI	(-0.104,-0.084)	(0.03,0.05)	(0.13,0.16)	(-0.076,-0.039)	(-0.047,-0.017)	(0.005,0.041)
95% CI (in SD)	(-0.682,-0.548)	(0.278,0.47)	(0.522,0.64)	(-0.5,-0.256)	(-0.443,-0.161)	(0.019,0.165)
Observations	15,364	15,364	15,364	15,364	15,364	15,364
R-squared	0.788	0.812	0.896	0.848	0.855	0.926
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1163	.1084	.5682	.1163	.1084	.5682
Pre-2007 SD	.1530	.1067	.2495	.1530	.1067	.2495
Schools	1782	1782	1782	1782	1782	1782

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 56: 4th Grade ELA Test Scores, English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0359***	0.0633***	0.102***	0.0177***	0.000104	-0.00248
	(0.00260)	(0.00290)	(0.00431)	(0.00404)	(0.00344)	(0.00616)
did (in SD)	-0.421***	0.485***	0.535***	0.208***	0.001	-0.013
	(0.031)	(0.022)	(0.023)	(0.047)	(0.026)	(0.032)
95% CI	(-0.041,-0.031)	(0.058,0.069)	(0.094,0.11)	(0.01,0.026)	(-0.007,0.007)	(-0.015,0.01)
95% CI (in SD)	(-0.482,-0.36)	(0.442,0.528)	(0.49,0.58)	(0.116,0.3)	(-0.05,0.052)	(-0.076,0.05)
Observations	21,607	21,607	21,607	21,607	21,607	21,607
R-squared	0.771	0.779	0.879	0.852	0.858	0.913
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0651	.1563	.6636	.0651	.1563	.6636
Pre-2007 SD	.0852	.1306	.1908	.0852	.1306	.1908
Schools	2692	2692	2692	2692	2692	2692

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 57: 4th Grade ELA Test Scores, Not English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.149***	0.0260***	0.157***	0.00341	0.00570**	0.0376*
	(0.0132)	(0.00486)	(0.0149)	(0.0168)	(0.00270)	(0.0220)
did (in SD)	-0.617***	0.436***	0.693***	0.014	0.095**	0.166*
	(0.055)	(0.081)	(0.066)	(0.070)	(0.045)	(0.097)
95% CI	(-0.175,-0.123)	(0.016,0.036)	(0.128,0.186)	(-0.03,0.036)	(0,0.011)	(-0.006,0.081)
95% CI (in SD)	(-0.725,-0.509)	(0.277,0.595)	(0.564,0.822)	(-0.123,0.151)	(0.007,0.183)	(-0.024,0.356)
Observations	7,416	7,416	7,416	7,416	7,416	7,416
R-squared	0.758	0.273	0.661	0.822	0.519	0.736
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.3579	.0121	.1937	.3579	.0121	.1937
Pre-2007 SD	.2414	.0597	.2266	.2414	.0597	.2266
Schools	1237	1237	1237	1237	1237	1237

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 58: 8th Grade ELA Test Scores, English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0249***	0.0102***	0.0241***	-0.00540	-0.000981	0.00168
	(0.00323)	(0.00319)	(0.00692)	(0.00571)	(0.00353)	(0.00768)
did (in SD)	-0.226***	0.130***	0.117***	-0.049	-0.012	0.008
	(0.029)	(0.041)	(0.033)	(0.052)	(0.045)	(0.037)
95% CI	(-0.031,-0.019)	(0.004,0.016)	(0.011,0.038)	(-0.017,0.006)	(-0.008,0.006)	(-0.013,0.017)
95% CI (in SD)	(-0.283,-0.169)	(0.05,0.21)	(0.052,0.182)	(-0.151,0.053)	(-0.1,0.076)	(-0.065,0.081)
Observations	11,949	11,949	11,949	11,949	11,949	11,949
R-squared	0.801	0.760	0.909	0.867	0.827	0.937
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0819	.0770	.4730	.0819	.0770	.4730
Pre-2007 SD	.1103	.0785	.2068	.1103	.0785	.2068
Schools	1737	1737	1737	1737	1737	1737

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 59: 8th Grade ELA Test Scores, Not English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.152***	0.00378**	0.0559***	-0.0424	0.000313	0.0224*
	(0.0184)	(0.00181)	(0.0108)	(0.0261)	(0.00104)	(0.0135)
did (in SD)	-0.686***	0.171**	0.526***	-0.191	0.014	0.211*
	(0.083)	(0.082)	(0.102)	(0.118)	(0.047)	(0.127)
95% CI	(-0.188,-0.116)	(0,0.007)	(0.035,0.077)	(-0.094,0.009)	(-0.002,0.002)	(-0.004,0.049)
95% CI (in SD)	(-0.849,-0.523)	(0.01,0.332)	(0.326,0.726)	(-0.422,0.04)	(-0.078,0.106)	(-0.038,0.46)
Observations	4,224	4,224	4,224	4,224	4,224	4,224
R-squared	0.787	0.216	0.480	0.834	0.401	0.599
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.4295	.0033	.0574	.4295	.0033	.0574
Pre-2007 SD	.2217	.0221	.1062	.2217	.0221	.1062
Schools	769	769	769	769	769	769

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 60: 4th Grade Math Test Scores, English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.0401***	0.107***	0.120***	0.00829*	0.0279***	0.00450
	(0.00262)	(0.00433)	(0.00462)	(0.00438)	(0.00604)	(0.00584)
did (in SD)	-0.476***	0.621***	0.658***	0.098*	0.162***	0.025
	(0.031)	(0.025)	(0.025)	(0.052)	(0.035)	(0.032)
95% CI	(-0.045,-0.035)	(0.099,0.115)	(0.111,0.129)	(0,0.017)	(0.016,0.04)	(-0.007,0.016)
95% CI (in SD)	(-0.537,-0.415)	(0.572,0.67)	(0.609,0.707)	(-0.004,0.2)	(0.093,0.231)	(-0.038,0.088)
Observations	21,561	21,561	21,561	21,561	21,561	21,561
R-squared	0.731	0.790	0.853	0.828	0.849	0.900
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.0527	.2671	.7709	.0527	.2671	.7709
Pre-2007 SD	.0843	.1722	.1824	.0843	.1722	.1824
Schools	2692	2692	2692	2692	2692	2692

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 61: 4th Grade Math Test Scores, Not English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.142***	0.0731***	0.191***	0.0480***	0.0513***	0.0650***
	(0.00901)	(0.00584)	(0.0117)	(0.0146)	(0.00919)	(0.0177)
did (in SD)	-0.709***	0.562***	0.713***	0.240***	0.395***	0.243***
	(0.045)	(0.045)	(0.044)	(0.073)	(0.071)	(0.066)
95% CI	(-0.16,-0.124)	(0.062,0.085)	(0.168,0.214)	(0.019,0.077)	(0.033,0.069)	(0.03,0.1)
95% CI (in SD)	(-0.797,-0.621)	(0.474,0.65)	(0.627,0.799)	(0.097,0.383)	(0.256,0.534)	(0.114,0.372)
Observations	9,443	9,443	9,443	9,443	9,443	9,443
R-squared	0.724	0.598	0.748	0.807	0.670	0.802
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.2164	.0766	.4515	.2164	.0766	.4515
Pre-2007 SD	.2003	.1300	.2679	.2003	.1300	.2679
Schools	1325	1325	1325	1325	1325	1325

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 62: 8th Grade Math Test Scores, English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.102***	0.0416***	0.130***	-0.0257***	-0.00545	0.0119
	(0.00524)	(0.00653)	(0.00778)	(0.00951)	(0.00786)	(0.00961)
did (in SD)	-0.585***	0.442***	0.549***	-0.147***	-0.058	0.050
	(0.030)	(0.069)	(0.033)	(0.054)	(0.084)	(0.041)
95% CI	(-0.112,-0.092)	(0.029,0.054)	(0.115,0.145)	(-0.044,-0.007)	(-0.021,0.01)	(-0.007,0.031)
95% CI (in SD)	(-0.644,-0.526)	(0.307,0.577)	(0.484,0.614)	(-0.253,-0.041)	(-0.223,0.107)	(-0.03,0.13)
Observations	11,898	11,898	11,898	11,898	11,898	11,898
R-squared	0.824	0.824	0.893	0.877	0.877	0.928
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.1649	.0873	.5029	.1649	.0873	.5029
Pre-2007 SD	.1745	.0941	.2368	.1745	.0941	.2368
Schools	1728	1728	1728	1728	1728	1728

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 63: 8th Grade Math Test Scores, Not English Proficient

VARIABLES	(1) Level 1	(2) Level 4	(3) Level 34	(4) Level 1	(5) Level 4	(6) Level 34
did	-0.144***	0.0365***	0.116***	0.0289	0.0200**	0.0559**
	(0.0175)	(0.00608)	(0.0165)	(0.0273)	(0.00794)	(0.0262)
did (in SD)	-0.567***	0.421***	0.509***	0.114	0.231**	0.245**
	(0.069)	(0.070)	(0.072)	(0.108)	(0.092)	(0.115)
95% CI	(-0.178,-0.11)	(0.025,0.048)	(0.084,0.148)	(-0.025,0.082)	(0.004,0.036)	(0.005,0.107)
95% CI (in SD)	(-0.702,-0.432)	(0.284,0.558)	(0.368,0.65)	(-0.098,0.326)	(0.051,0.411)	(0.02,0.47)
Observations	5,038	5,038	5,038	5,038	5,038	5,038
R-squared	0.781	0.651	0.765	0.836	0.735	0.812
Year Fixed Effects	X	X	X	X	X	X
School Fixed Effects	X	X	X	X	X	X
Time trends				X	X	X
Pre-2007 Mean	.4473	.0311	.2171	.4473	.0311	.2171
Pre-2007 SD	.2538	.0866	.2277	.2538	.0866	.2277
Schools	833	833	833	833	833	833

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 64: Graduation Rates: Special Education and General Education

VARIABLES	(1) Graduation Rate (Disabled)	(2) Graduation Rate (Disabled)	(3) Graduation Rate (Gen Ed)	(4) Graduation Rate (Gen Ed)
did	-0.0398***	-0.124***	0.0915***	0.0604***
	(0.0136)	(0.0167)	(0.00781)	(0.00769)
did (in SD)	-0.127***	-0.395***	0.272***	0.180***
	(0.043)	(0.053)	(0.023)	(0.023)
95% CI	(-0.066,-0.013)	(-0.157,-0.091)	(0.076,0.107)	(0.045,0.075)
95% CI (in SD)	(-0.211,-0.043)	(-0.499,-0.291)	(0.227,0.317)	(0.135,0.225)
Observations	10,915	10,915	12,130	12,130
R-squared	0.731	0.799	0.877	0.912
Year Fixed Effects	X	X	X	X
School Fixed Effects	X	X	X	X
Time trends		X		X
Pre-2007 Mean	.5702	.5702	.6915	.6915
Pre-2007 SD	.3140	.3140	.3359	.3359
Schools	1716	1716	2026	2026

Standard errors clustered by school in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Number of observations is reported as the total number of students for all post-treatment years from 2008 to 2015. Number of schools is reported for the first year of treatment, the 2007-2008 school year. Dependent variables are the NY state ELA and math scores, both in raw percentage point changes and standardized to have a mean of zero and a standard deviation of one. Scores from all 3 years of implementation are used. Two regressions are presented: one controlling for year and school fixed effects, and another controlling for year fixed effects, school fixed effects, and school-specific linear time trends.

Table 65: Propensity Score Matching, Attendance, Suspension, and Graduation Rates

VARIABLES	(1) Attendance	(2) Suspensions	(3) Graduation
did	0.00430	-0.000580	0.207***
	(0.00387)	(0.00436)	(0.0128)
Observations	4,222	4,219	1,145

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 66: Propensity Score Matching, Aspirations

VARIABLES	(1) PER_2YR_ COLLEGE	(2) PER_4YR_ COLLEGE	(3) PER_ EMPLOYMENT	(4) PER_POST_ SECONDARY	(5) PER_ MILITARY
did	-0.0306	0.112***	0.00421	0.0138***	0.00883***
	(0.0268)	(0.0314)	(0.00759)	(0.00529)	(0.00282)
Observations	461	303	263	191	177

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 67: Propensity Score Matching, Test Scores

VARIABLES	(1) LEVEL1__TESTED	(2) LEVEL4__TESTED	(3) LEVEL34__TESTED
ELA4	-0.0610***	0.00282	0.0706***
	(0.00618)	(0.0133)	(0.00777)
ELA8	-0.0464**	0.0103***	0.0328**
	(0.0223)	(0.00319)	(0.0151)
Math4	-0.0663***	0.104***	0.106***
	(0.00635)	(0.0108)	(0.00784)
Math8	-0.0461***	0.0428***	0.0881**
	(0.0167)	(0.0113)	(0.0373)
Observations	2,309	2,309	2,309

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Propensity scores generated by performing a logit regression of the NYC dummy variable on total school enrollment and school-wide percentages of students receiving free lunch, students receiving reduced lunch, students with limited English proficiency, white students, black students, Hispanic students, Native American students, and Asian students. Pairs of schools, one inside and one outside NYC, are matched based with a caliper of 0.07.

Table 68: DFL Re-Weighting Coefficients

VARIABLES	(1) Standard	(2) DFL Re-Weighting
Attendance Rate	0.00460	0.00449
Suspensions	0.02668	0.02671
Graduation Rates	0.11367	0.12315
Two-Year College	0.01573	0.01534
Four-Year College	0.12545	0.12496
Employment	0.01301	0.01296
Trade School/Other	0.00245	0.00260
Military	-0.00236	-0.00235
ELA4 Level 1	0.00187	0.00134
ELA4 Level 4	0.00700	0.00733
ELA4 Passing	-0.02126	-0.02119
ELA8 Level 1	-0.00221	-0.00267
ELA8 Level 4	0.00081	0.00087
ELA8 Passing	-0.04520	-0.04608
Math4 Level 1	-0.03147	-0.03240
Math4 Level 4	0.03613	0.03710
Math4 Passing	0.04535	0.04683
Math8 Level 1	-0.06007	-0.06117
Math8 Level 4	-0.02633	-0.02671
Math8 Passing	-0.01036	-0.01096

Note: Effects are in diff-in-diff percentage point changes over time

Table 69: Effects of Other Selected Policy Interventions on Test Scores

Location	Treatment	Variable	Effect (in SD)
NYC	FSF	4th Grade ELA Test Scores	-0.092*** (0.022)
NYC	FSF	4th Grade Math Test Scores	0.214*** (0.025)
NYC	Up to $250 payment for test scores (Allan and Fryer 2011)	4th Grade ELA Test Scores	-0.026 (0.034)
NYC	Up to $250 payment for test scores (Allan and Fryer 2011)	4th Grade Math Test Scores	0.062 (0.047)
NYC	Teacher Incentives (Fryer 2013)	4th Grade ELA Test Scores	-0.020 (0.014)
NYC	Teacher Incentives (Fryer 2013)	4th Grade Math Test Scores	-0.033 (0.017)

Note: Units are schools in the study. FSF effects represent changes in passing scores in linear time trends model. Effects have been standardized to have a mean of zero and standard deviation of one. Therefore, these units are in SD units. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.