Educational Resources and Student Performance:

Evidence from the Save Harmless Provision in New York State

Philip Gigliotti and Lucy C. Sorensen

University at Albany, State University of New York

Abstract: A long-standing debate in the economics of education literature is whether increasing educational resources moves the needle on student achievement. Education finance reformers advocate delivering extra resources to disadvantaged school districts to close academic achievement gaps, but their efforts are subject to criticism from skeptics who believe that extra resources do not actually improve performance. This study leverages variation in per-pupil expenditures from a specific provision of the state aid formula in New York State that allows districts to maintain prior levels of total state aid even as their student enrollment declines. We uncover performance gains of approximately .04 standard deviations in both math and English corresponding to $1,000 in additional per-pupil spending, with effects driven primarily by reductions in the ratio of students to teachers. This study strengthens the case that school resources matter, and that targeted financial investments can help close educational achievement gaps. **JEL Codes** I22, C36, H75.

1 Introduction

Disparities between disadvantaged students and their wealthy counterparts are a regular empirical finding in the education literature. Racial minorities have on average lower scores on standardized achievement tests and lower graduation rates (Fryer and Levitt 2006; Hanushek and Rivkin 2006; Heckman and LaFontaine 2010). Research suggests that much of this gap reflects underlying socioeconomic differences (Clotfelter, Ladd and Vigdor 2009; Fryer and Levitt 2004). Today, individuals in the lowest income decile have four years less educational attainment than individuals in the highest income decile, at a time when education has become even more essential to financial stability. Between 1997 and 2007, wages grew by 25% for college graduates, while they stagnated for high school graduates and declined by 13% for high school dropouts (Duncan and Murnane 2011). If education and income are causally linked, educational achievement gaps will lead to widening socioeconomic disparities and income inequality.

While consensus exists on the presence and consequences of academic achievement gaps, solutions remain more controversial. The first series of U.S. education finance reforms focused on equalizing educational expenditures between districts, and the second series then attempted to deliver supplementary resources to low-performing districts to account for high need student populations. But while states nationwide have been largely successful in these instrumental goals, disparities in performance persist (Lafortune, Rothstein and Schanzenback 2016; Yinger 2004). Hanushek (1994) documents 3.5% real annual increases in per-pupil expenditures (PPE) between 1970 and 1990, and Bifulco (2005) documents that, since 1987, PPE in the average black student’s district have outpaced those in the average white student’s district by approximately $400. Despite these massive investments, there have been few obvious improvements in disadvantaged public school districts, and a complicated empirical history has left researchers unconfident in claiming a definitive causal link between educational expenditures and student performance.

Dating back to early attempts in the 1960's (Coleman, 1966) scholars struggled to link educational resources to a definitive positive impact on student achievement. In the late 1990's, two teams conducted a famous pair of metanalyses on the topic, with Hanushek (1997) claiming no relationship and Greenwald, Hedges and D. (1996) claiming a positive relationship. As this history pre-dated the widespread adoption of quasi-experimental methods in economics research, it is now clear that the early body of research failed to identify the nature of this relationship. Jackson, Johnson and Persico (2016) suggest that direct estimates of the effect of educational resources on student performance are likely to be biased downwards, since educational policymakers often invest additional resources to low performing schools. This phenomenon creates a simultaneity issue that cannot be directly controlled for and which must be addressed through quasi-experimental methods.

More recent research, focusing mostly on the impacts of state level school finance reforms has found evidence that is suggestive of a positive relationship between educational resources and student achievement which develops over time (Lafortune et. al., 2016; Card & Payne, 2002; Guryan, 2001). An authoritative study on this subject by Jackson et al (2016), measures the effect of per pupil educational expenditures on long term student outcomes and finds a positive effect that is stronger for disadvantaged students. However, very few studies have identified the effect of current year per-pupil educational dollars on immediate academic outcomes. Only Papke (2005) delivers such an estimate, using exogenous variation in state educational aid resulting from Michigan's school finance reform to identify positive impacts on student outcomes, and finding that a 10% increase in current year per-pupil expenditures is associated with a 2 percentage point increase in the pass rate on year end examinations. A parameter estimate of the immediate impact of each dollar of educational spending is perhaps the most salient measure of the impact of educational resources for educators and policy-makers seeking to demonstrate results and meet accountability targets. Therefore, new and well-identified estimates of this effect should be of intense interest to those in the educational policy community.

This paper contributes to the literature by employing quasi-experimental methods to investigate the relationship between current year per-pupil educational expenditures and student-teacher ratios and immediate student performance in New York State (NYS) school districts between the 2007-08 through 2014-15 academic years. It relies on variation in district level state aid to identify this relationship. During the 2007-08 school year, NYS reformed its education finance system and implemented a need-based foundation aid formula that included a number of idiosyncratic rules and policies. One of these provisions, called “Save Harmless,” stipulated that districts could not lose money if their estimated need declined. The largest impact of this provision was that districts did not lose funding when their enrollment decreased, leading districts with declining enrollment to have systematically higher per-pupil expenditures. While this policy was in place, New York experienced the highest levels of population loss in the country, causing extreme changes in enrollment and rapidly compounding increases in educational resources. We demonstrate that enrollment change was uncorrelated with most demographic characteristics that would indicate a change in district composition. The confluence of these factors creates a natural experiment, and produces plausibly exogenous variation in school resources.

Leveraging this variation through instrumental variable estimation, we find positive effects of expenditures on elementary and middle school test scores. Our estimates are somewhat large in comparison to those of Lafortune et. al. (2016). We probe the validity of these inferences, specifically the validity of our exclusion restriction assumptions, through a series of robustness checks assessing key threats to validity, and find little evidence that they bias our results. We conclude that the Save Harmless policy treatment provides quasi-exogenous variation in school resources, conditional on district and year fixed effects, district enrollment and district level demographic composition.

This study is the first to our knowledge to assess the impact of a Save Harmless provision on student achievement. In addition, it is the first to generate causal estimates of the impacts of educational resources in New York State following their 2007 school finance reform, and the most recent estimate of the impact of current year educational resources on immediate academic outcomes. Since the New York reform has been noteworthy to scholars for both the magnitude of its investments, and the political controversies surrounding it, identification of the effects of these investments can inform arguments about optimal levels of spending and possible adjustments to state aid formulas. Furthermore, we assert that these findings likely generalize to state-level education finance reforms nationwide, demonstrating that states can realize meaningful gains in student achievement when they make large investments in their public school systems.

2.1 Background on School Finance Reforms

Traditionally districts serving poor students have weaker property tax bases and therefore less revenue available per pupil. These funding gaps between socioeconomically disadvantaged and wealthy public school districts have acted as a common target for educational reforms. Such efforts led to state-level school finance reforms (SFR's) beginning in that 1970's, which sought to equalize spending across districts, and adequacy-based SFR's beginning in the 1980's and 1990's which delivered extra resources to low-performing districts. Hanushek (1994) charts the trajectory of the early equity-based finance reforms, documenting 3.5% annual increases in expenditures between 1970 and 1990. Lafortune, Rothstein and Schanzenback (2016) analyze later adequacy-based reforms and document a 40% increase in spending between 1990 and 2012, which was concentrated in low-performing districts.

While an extensive literature examines the impacts of SFR's on funding levels, research into their impacts on student performance has been less comprehensive. Card and Payne (2002) analyze a national sample of pre-1992 data and suggest that SFR's led to reduction in achievement gaps between rich and poor students. Guryan (2001) found mixed evidence that SFR's improved test scores in Massachusetts. Lafortune, Rothstein and Schanzenback (2016) analyze post-1990 reforms and found effects on student achievement that develop incrementally over time. Jackson, Johnson, and Persico (2016), summarized in more detail below, looked broadly at historical school finance reforms and long-term student outcomes using an event study design. Nonetheless, the literature on student outcomes following school finance reforms is relatively sparse, and while suggestive of positive impacts, does not lend itself to definitive conclusions.

Further spending in high-poverty schools should be justified by substantial empirical evidence that expenditures produce a causal impact on student achievement. An extensive literature has attempted to document the relationship between expenditures and student outcomes, however researchers disagree over whether the evidence conclusively demonstrates that this connection exists. In the late 1990's, two independent research teams analyzed the existing literature, and arrived at different conclusions. Hanushek (1997) analyzed 90 studies and found that only 27% of studies estimate a positive and statistically significant coefficient on educational expenditures, concluding no positive relationship between spending and performance. Greenwald, Hedges and D. (1996) arrive at different conclusions than Hanushek (1997) despite a similar meta-analytic approach. In a sample of 60 district or school level studies, they find positive and statistically significant coefficients on expenditures in 44% of their sample, and perform combined significance tests leading them to conclude that spending has a meaningful effect on student achievement. Verstegen and King (1998) also argue forcefully for the efficacy of educational expenditures. After reviewing 35 years of research, they claim conclusive evidence that factors such as class size and teacher quality improve student outcomes, and since these factors raise the cost of instruction, expenditures are linked in a causal chain to student achievement. High quality experimental and quasi-experimental research has shown that reducing class size increases student achievement (Angrist and Lavy 1999; Krueger 1999), but some have shown null effects (Hoxby 2000).

These meta-analyses largely contained studies employing education production function designs with endogenous operationalization of school resources. Due to these endogeneity concerns it is unlikely that in aggregate these studies could identify unbiased estimates of the effect of spending on student performance. Because education policy-makers commonly deliver extra resources to low-performing schools or cohorts of students with higher need, direct estimates of this relationship, even with district fixed effects, will likely be biased downwards. More sophisticated contemporary research has used quasi-experimental methods to estimate effects of educational expenditures on student outcomes. Using state aid reforms as an instrument for educational expenditures, Jackson, Johnson and Persico (2016) identify causal relationships between per-pupil expenditures (PPE) and completed schooling, wages and reduced adult poverty in a national sample. Furthermore they find that PPE have a larger effect on performance in socioeconomically disadvantaged student populations. Papke (2005); Papke (2008) and Chaudhary (2009) use state aid grants as an instrument for resources, leveraging variation resulting from state aid reforms in Michigan. While other studies (Card and Payne, 2002; Guryan, 2001; and Lafortune et al., 2016) suggest positive impacts of SFR's, and by extension educational resources, on student performance, to date, these are the only studies to our knowledge that use methods beyond district and year fixed effects to discern effects of per-pupil financial resources on student performance in the U.S. context. Furthermore, only Papke (2005) identifies a parameter estimate for the effect of current year per-pupil educational expenditures on student performance, and no similar estimate has been produced since 2009.

This study will contribute to the literature by developing a novel approach for identifying effects of increased per-pupil expenditures on student achievement. Specifically, by leveraging variation in funding from the "Save Harmless" policy in the NYS state aid formula, we will show it is possible to derive unbiased estimates of the impact of current year per-pupil expenditures on student outcomes. This natural experiment differs from previous quasi-experimental approaches, which studied the impact of large scale investments of funding resulting from school finance reforms. Unlike those studies, which estimated the effect of increasing resources, this study explores the effect of holding resources constant and distributing them to a smaller group of students. This research will contribute to ongoing policy debates surrounding educational finance reform in New York State, and can generalize to nationwide debates over education finance reforms and the role of resource inputs in educational production functions.

2.2 New York State Aid Reform and the "Save Harmless" Provision

Over the past quarter century, New York State has been a hotbed of school finance reform. In the mid 1990's, despite average property tax rates that were among the highest in the nation, less-privileged schools demonstrated persistently substandard performance. According to reformers, state aid programs that had focused on equalizing spending between districts failed to consider substantial cost differences between districts. An extended debate over optimal solutions led to several proposals for new state aid funding strategies. Leading this effort, Duncombe and Yinger (2000) argued for a performance-based formula that would account for student characteristics such as poverty and limited English proficiency, as well as for regional cost disparities. Such a formula would attempt to go beyond equalizing spending and invest more resources in disadvantaged school districts to equalize performance.

Educational reform in NYS was not only an academic exercise, but was the target of intense political advocacy. In 1993, an advocacy group known as the Campaign for Fiscal Equity launched a protracted legal campaign to deliver financial remediation to underperforming schools. Their advocacy led to years of litigation, culminating in a landmark ruling in the case of Campaign for Fiscal Equity vs. the State of New York (2003). The court declared that the state of New York had violated the constitutional right of students to a sound basic education, and directed the state to implement educational finance reforms to close performance gaps between school districts. This resulted in the Education Budget and Reform Act of 2007 (EBRA) which introduced a performance-based foundation aid formula that adjusted for pupil needs, including enrollment, poverty and limited English proficiency, along with regional cost differences (Abbott, Hodgens and Wenzel 2013).

One noteworthy provision of the new foundation aid formula was that it maintained an archaic provision of NYS education finance called the "Save Harmless" provision. Under this clause, school districts that experienced declining enrollment would not experience cuts to their state aid allotment. Save Harmless was implemented in 1976, to prevent schools with declining enrollment from shutting down (Levine 1976). The policy was immediately met with criticism, as many worried that it privileged certain school districts at the expense of others. As early as the 1980's, this provision was a political football being debated in relation to equity issues. For example, a 1983 New York Times article described debate over whether wealthy school districts should receive Save Harmless guarantees (Chira 1983). When the EBRA was implemented at the start of the 2007-08 academic year it maintained the Save Harmless provision, guaranteeing that districts with declining enrollment or pupil need would not only receive equal funding to the year prior, but would also receive a 3% adjustment for inflation. This provision drew the ire of some education reformers, who claimed that the provision delivered millions of dollars in aid to students who didn't exist, when those resources could have been distributed to high-need districts in pursuit of equity gains (Cunningham 2014).

The Save Harmless provision offers the potential for a natural experiment, and we harness the variation in state educational aid to school districts arising from the policy. Under this provision, districts that experience declining enrollment receive artificially inflated levels of per-pupil educational aid. Employing a conditional exogeneity argument, we demonstrate that, conditional on district and year fixed effects and current student enrollment, within-district year-to-year enrollment change and the corresponding impact of the Save Harmless provision generate plausibly exogenous variation in educational resources. We defend this argument through a rigorous set of robustness checks evaluating the key threats to validity.

2.3 Save Harmless and NYS Demographic Trends

The execution of the Save Harmless policy is especially salient in NYS, due to demographic trends occurring during the period of our study. At the turn of the new millennium, NYS was experiencing steady growth, but analysts noted declining population in the "Rust Belt" cities of Upstate NY and upstate counties in general (Wing 2003). Over the next ten years growth declined, with NYS's percentage growth rate ranking 46th in the nation and growth concentrated in the downstate region with 17 upstate counties losing population. (NYS Department of Labor 2011). Between 2010 and 2015, the last five years of our study, this trend multiplied, with 41 out of 50 upstate counties losing population. While downstate growth was slow at a paltry .33%, upstate NY lost 65,638 people for a growth rate of -1.04%, producing a statewide growth rate of -.12% (Empire Center 2016). In 2016, Forbes ranked NYS number 1 for losing the most net migrants nationwide, with 126,000 people leaving the state (Kotkin 2016).

The declining statewide population carried over to declining enrollment in NYS school districts. Over the period in our sample, the mean percentage change in enrollment from the 2006-07 academic year to the 2014-15 academic year was -5.61% with 81% of observations in our sample showing declining enrollment. The mean percentage change in observations with declining enrollment was -8.17%. Figure I first shows graphically how the rate of enrollment change between 2007 and 2015 varies across school districts in the state. Out of 652 districts in our analysis sample, only 42 experienced zero years of declining enrollment during the period of our study. The second map in Figure I indicates how, despite consistent enrollment declines in the upstate region, total inflation-adjusted expenditures still increased over time. These two patterns in juxtaposition, create the growth in per-pupil expenditures we investigate in this study.

The graphics in Figure II further demonstrate that enrollment declines were a persistent trend in NYS school districts and led to marked increases in per-pupil expenditures (PPE). All graphics absorb district fixed effects. The plot of enrollment over time shows that enrollment declined year after year in districts across NYS, starting at a peak of 2,650 in the 2007-08 academic year and falling to 2,450 in 2014-15. The plot of the relationship between PPE and percent enrollment decrease demonstrates the impact of the Save Harmless policy. As enrollment declines, district PPE increases markedly, with the graphic showing a steep trend. This illustrates the logic behind our natural experiment; districts that experienced declining enrollment had clear increases in PPE, and enrollment declines seem consistent across district types, which should make the variation in PPE plausibly exogenous.

The biggest potential threat to this natural experiment is that district demographic composition changed as enrollment declined in a way which would influence student performance. This could occur if students from certain demographic groups were more likely to move out of a district. We explore these relationships and find that while three out of four demographic characteristics (percent of students with free lunch eligibility, percent of students with Limited English Proficiency, and percent of students with disabilities) have minimum association with our measure of enrollment change, percent minority enrollment is significantly correlated, indicating minority students are more likely to move out of the district. In robustness checks we probe the effect of demographic controls including percent minority composition on our estimates and find that they have minimal impact on effect size or significance. We conclude that student selection mechanisms should not seriously bias our findings.

3. Data

The analysis in this paper is based on publicly available data from the New York State Education Department (NYSED). Financial variables are drawn from the Fiscal Analysis and Research Unit's (FARU) Fiscal Profile Reporting System (FPRS). Test scores, demographics and other control variables are drawn from NYSED School Report Card data, which contain both school- and district-level data. Merging these data sources together results in a complete set of 678 district matches accounting for all districts in New York State included in the FARU data. Since some of the districts only serve elementary or high school students, but not both, we further restrict our sample to only K-12 school districts, resulting in a sample of 652 districts. The sample excludes the New York City Public School District, because it differs meaningfully from the rest of the state on size, population, and organizational and financial structure. (The data on NYC schools was furthermore incompatible for matching to the fiscal database.) Data spans an eight-year period from the 2007-08 to the 2014-15 academic years.

The key explanatory variable of interest in this study is *Save Harmless*, a continuous measure of the inverse percentage change in enrollment in each district from the 2006-07 academic year. This serves as a proxy variable for the effect of the Save Harmless provision since districts experiencing larger enrollment declines would receive larger benefits in terms of available per-pupil foundation aid. It is calculated by first subtracting enrollment in the current period from base year enrollment, and dividing by enrollment in the base year. If there is a decline in enrollment from the base period, this variable will have a positive value. To aid interpretation we multiply this ratio by 100. Finally, we code it as 0 for all years with enrollment increases from the base year. This ensures that we are only measuring impacts on spending resulting from enrollment decrease and the associated Save Harmless provision.

Because this variable is dependent on student enrollment, and enrollment can be measured in different ways, selection of an appropriate measure is essential to successful identification in our models. We choose a measure from the FARU data set called Duplicated Combined Adjusted Average Daily Membership (DCAADM). This measure includes not only students enrolled in district classes, but all students for which the district has financial obligations, including charter school students, students in alternative special education placements, incarcerated students, and students for which the district pays tuition to another district, such as in urban-suburban programs. It is important to include all students for which the district has a financial obligation when working with per-pupil financial measures, because failing to include these could misspecify the fiscal impact of declining enrollment, biasing resulting estimates. NYS considers DCAADM the best enrollment measure to use when working with financial variables, and calculates its official per pupil financial statistics using this measure

The dependent and endogenous variables in this study include both institutional and student outcomes. The institutional variables include a measure of expenditures per pupil (total expenditures divided by DCAADM), and a measure of the student-teacher ratio which is BEDS enrollment (a measure of enrollment in district classes) divided by total teachers. We also test state aid revenues per pupil (total state aid divided by DCAADM), total instructional expenditures, and the following expenditure ratios: instructional (instructional expenditures divided by total expenditures), debt (debt payments divided by total expenditures), administrative (administrative expenditures divided by total expenditures) and teacher retirement (teacher retirement expenditures divided by total expenditures. The student outcomes are average mathematics and English standardized tests scores for grades three through eight. This is an averaged measure of district performance on end of year standardized testing administered to all students in grades three through eight. The test score scale changed in the 2012-13 academic year, and so we standardize these measures by year to have mean of zero and standard deviation of one.

Finally, our analysis considers district-level student demographic characteristics. These include percent free lunch eligibility, percent racial or ethnic minority, percent limited English proficiency (LEP), and percent of students with disabilities. Two of these variables were calculated by the authors, and two were directly reported from New York State. Percent minority is calculated by subtracting percent white from 100. Percent students with disabilities is calculated by dividing total students with disabilities in grades three through eight by total students in grades three through eight. Analysis in our models employs data from the 2007-08 academic year, in which the EBRA was implemented, to the 2014-15 academic year. All financial variables are adjusted for inflation, reported in year 2016 dollars and divided by 1,000. Additional minor variable construction details are provided in an appendix. We provide summary statistics of all dependent and independent variables in Table I.

4 Methods

Our analysis has two primary objectives. First, we use a two-way fixed effects model to estimate the direct effects of year-to-year enrollment declines on per-pupil expenditures and student-teacher ratio. Second, we use an instrumental variables strategy to estimate the impact of per-pupil expenditures and student-teacher ratios on student performance, assigning the Save Harmless policy as an instrument for expenditures. These primary methods, along with a comprehensive set of robustness tests, are described in this section.

4.1 Estimating the Effect of Enrollment Declines on Institutional Resources

To evaluate the direct impact of the Save Harmless policy on institutional outcomes we estimate a two-way fixed effects model according to the following equation:

$$y_{dt} = \gamma_0 Save\ Harmless_{dt} + \gamma_1 Enrollment_{dt} + \gamma_2 X_{dy} + \theta_d + \tau_t + \varepsilon_{dt} \quad \text{(Equation 1)}$$

Where y is an outcome of interest for district d in year t, γ_0 is the coefficient on our primary explanatory variable $Save\ Harmless$ (representing the degree of enrollment decline) and γ_1 is the effect of enrollment (DCAADM), controlling for which allows us to isolate the impact of the Save Harmless conditional on the level of student enrollment. γ_2 represents coefficients on X, a vector of district level time-varying characteristics, and ε is a stochastic error term for district d in year t. We include θ_d, a vector of district fixed effects which can control for time-invariant characteristics of the district, including any underlying factors that may cause the district to have larger or smaller average student enrollment declines. We also include a vector of year fixed effects τ_t which accounts for all observed and unobserved characteristics of each year in our sample, including secular time trends in population change. All models are estimated with Huber-White robust standard errors clustered by district, to address heteroscedasticity and autocorrelation within districts.

4.2a Estimating the Effect of Per-Pupil Expenditures on Student Performance

A more generalizable question than the impact of the Save Harmless policy on institutional outcomes, is the direct effect of per-pupil educational expenditures (PPE) and student-teacher ratios on student performance, which constitutes a perennial debate in the educational literature. A naïve approach to estimating this relationship would be to estimate an ordinary least squares model of the following form:

$$y_{dt} = \alpha_0 + \alpha_1 PPE_{dt} + \alpha_2 X_{dt} + \varepsilon_{dt} \quad \text{(Equation 2)}$$

Where y is an outcome of interest for district d in year t, α_0 is the y-intercept, α_1 is the association of PPE with educational performance, α_2 is a vector of coefficients on a vector of district level demographic and institutional characteristics (X), and ε is a stochastic error term for district d in year t . However, estimates of this form are likely to contain bias, since the level of school resources in a district is correlated with characteristics of the student population. Even after incorporating district and year fixed effects, estimates are likely to be downward biased because US educational authorities typically target increased resources to districts with current performance deficits or higher need (Jackson et al., 2016). This endogeneity violates the necessary assumptions to identify an unbiased estimate of the coefficient α_1 through linear regression.

If *Save Harmless* has a strong independent impact on PPE and student-teacher ratios, as we have predicted, it is possible to use our models from section 4.1 as a first stage in a two-stage least squares (2SLS) model to estimate the direct impact of PPE on student performance. This requires us to either demonstrate that the policy is unlikely to affect school districts in any way other than through its direct impact on educational resources, or to fully control for these alternate mechanisms. If we can isolate the exogenous impact of *Save Harmless*, we can use the two-way fixed effect models to predict values of district PPE, and then use predicted values in a second model to identify the effect of educational expenditures on student outcomes.

We describe our 2SLS approach in the following two equations. First, we estimate the first stage model according to Equation 3.

$$Resources_{dt} = \beta_1 Save\ Harmless_{dt} + \ \beta_2 Enrollment_{dt} + \ \beta_3 X_{dt} + \theta_d + \tau_t + \varepsilon_{dt}$$

(Equation 3)

In this equation, $Resources$ is either expenditures per pupil or student-teacher ratio for district d in year t, β_1 is the impact of enrollment change under *Save Harmless* on per-pupil expenditures, β_2 is the effect of enrollment, β_3 contains coefficients on X, a vector of time-varying district characteristics which block alternative mechanisms by which *Save Harmless* could impact performance, θ_d is a vector of district fixed effects, τ_t is a vector of year fixed effects, and ε is a stochastic error term for district d in year t

If we accept the conditional exogeneity argument – which we defend in the following section – then these predictions of expenditures per pupil are un-confounded by unobservable impacts of *Save Harmless.* We can use them to identify the impact of PPE and student-teacher ratio on performance in a second stage equation of the following form:

$$Student\ Outcome_{dt} = \delta_1 \widehat{Resources}_{dt} + \delta_2 Enrollment_{dt} + \delta_3 X_{dt} + \theta_d + \tau_t + \varepsilon_{dt}$$

(Equation 4)

Here, $Student\ Outcome$ is one of our four measures of student performance in district d in year t, δ_1 is the effect of a one thousand dollar increase or one student decrease in $\widehat{Resources}$, our predicted values of school resources. All other terms are identical to the first stage model. If our assumptions hold, then δ_1 will reflect the true impact of educational resources on student performance. All models are estimated with Huber-White robust standard errors clustered by district, to address heteroscedasticity and autocorrelation within districts.

4.2b Save Harmless Treatment as an Instrument

We show in the results section that *Save Harmless* is a strong predictor of PPE, but to serve as a plausible instrument, our variable must meet three additional assumptions. These are conditional independence, the exclusion restriction, and monotonicity. We interrogate these assumptions with a series of falsification tests. We only require the instrument to be exogenous conditional on the controls in our model, including: current student enrollment levels, district and year fixed effects, and the vector of time-varying student and institutional characteristics.

Conditional Independence

The first assumption that our instrument must meet is that of conditional independence. For this assumption to be upheld, our instrument must not be correlated with the error term in our outcome equation. The largest threat to this assumption is that the composition of test-taking students may change as enrollment declines, although it is unclear whether such a change would represent positive or negative selection out of school districts.

We conduct a test of the conditional independence assumption which regresses demographic characteristics of students on our measure of Save Harmless treatment, in a model similar to the first stage equation with district and year fixed effects (Table A.3). Save Harmless effects on several measures of student characteristics, including percent free lunch eligibility, percent students with disabilities and percent limited English proficiency, are insignificant. Only percent minority enrollment fails the test. A one percentage point decline in enrollment is associated with a 0.15 percentage point decrease in minority composition. We control directly for percent minority enrollment along with the other demographic variables in our models, so this finding poses no direct threat to validity, but could potentially signal the existence of some minor unobservable selection mechanism.

Exclusion Restriction

In order for our instrumental variables assumptions to hold, our instrument must not affect the academic performance measures through any mechanism other than educational resources. If there is another variable correlated with both the instrument and the outcome, it has the potential to confound our estimates. The finding that one of the four demographic covariates is correlated with our instrument does warrant some attention[1]. While it is possible that there is student sorting on unobservables, we believe that our controls for four major types of student demographics significantly mitigates these effects. In the robustness checks we probe the sensitivity of our results to these controls and conclude that sorting effects have very small impacts compared to the size of our estimates.

One further potential threat to the validity of our instrument is that effects could be driven by institutional impacts associated with enrollment change that are not caused by educational resources. One likely way this could happen were if our results were being driven by observations in which districts experienced enrollment increase. In these observations, districts receive a large influx of new students. A significant literature documents a negative impact on student achievement resulting from students changing schools. For instance, Rockoff and Lockwood find that movement from elementary to middle school is associated with a .15 standard deviation decline in test scores (.0015 per 1% increase), and Brummet (2014) documents that displacement from mass school closures in Michigan caused negative impacts of .002 standard deviations per 1 percentage point increase in displaced students within a school. If our results were driven by observations with enrollment increase we could be detecting these

[1] Since our instrument is uncorrelated with free lunch eligibility, disability status and limited English proficiency, it is possible that the minority variable doesn't carry any useful information about academic performance. We test this in our data by regressing our academic outcomes on the full vector of controls in our first stage model, excluding our instrument. While % minority eligibility has no impact on English test scores, it is significantly related to Math scores. A 1% increase in minority composition results in a .0075 standard deviation decrease in test scores. We compare this to reduced form estimates of our instrument's impact on Math scores. A 1% decrease in enrollment produces a .0086 increase in test scores, in models that control for minority composition. Since a 1% decrease in enrollment is associated with a .1562 % decrease in minority composition, uncontrolled estimates would be biased upwards by approximately .0012 standard deviations. These effects are small in comparison to our reduced form estimates, which are already corrected for this endogeneity. Since the confounder has no impact on English scores we consider these estimates fully robust.

effects. We consider this unlikely. Our measure of enrollment change is coded 0 for all observations with enrollment increase, which would dampen any effect from these observations. Nonetheless, we test this possibility in a robustness check and find the results largely confirm this assumption.

Monotonicity

Finally, for our instrumental variables assumptions to hold, the enrollment change measure must meet the assumption of monotonicity. This means that percentage change in enrollment ("Save Harmless") should affect PPE and student teacher ratio in the same direction in all districts. A simpler version of this test might be to demonstrate that enrollment decline has a positive relationship with PPE in subgroups of district types. We partition our sample based on four need to resource categories specified by NYS to classify districts by wealth and population concentration. These are high need urban, high need rural, medium need and low need. If the effect is positive or statistically indistinguishable from zero in all models, then the assumption is upheld. We run our first stage regression in the four subsamples of districts. The results are included in Table A.7 and A.8. The results show a positive relationship of enrollment decline with PPE and a negative relationship with student teacher ratio in all models. These findings illustrate that the impact of Save Harmless does not vary substantively across different district types.

5. Results

The empirical strategy in this study proceeds in two steps. First, we employ a two-way fixed effects model to identify the impacts of Save Harmless treatment on institutional outcomes in NY school districts. In the second step, we employ a two-stage least squares (2SLS) approach to identify the direct impacts of PPE and student-teacher ratio on student outcomes.

Our first set of results pertains to the two-way fixed effects models estimating the impact of the *Save Harmless* enrollment decline measure on educational inputs (per-pupil expenditures, state aid per-pupil and average class size as measured by the student-to-teacher ratio). The results of these models are provided in Table II. All models contain district and year fixed effects. Column 1 shows large and statistically significant effects of enrollment declines on PPE. This makes intuitive sense; if Save Harmless districts receive equal funding as prior years in which they had fewer students, dividing their total expenditures by a smaller enrollment figure will lead to larger measures of PPE. The average effect on treated districts is an increase of approximately $200 dollars per pupil per 1% decline in enrollment. This finding carries over to state aid. Column 2 shows that a 1% decline in enrollment predicts a $90 increase in state aid per-pupil. Column 3 shows a statistically significant decline in student teacher ratios of approximately .06 students per teacher corresponding to a 1% enrollment change. These results also confirm logical predictions; if Save Harmless districts lose students, but funding is sustained to avoid layoffs, we should see a decline in average student-teacher ratios.

Next, we examine the effect of Save Harmless on institutional spending decisions (Tables 3 and 4). We consider impacts on the four major spending categories: instructional spending, teacher retirement, debt payments and administrative expenses. Together these make up 98% of expenditures. The results show minor shifts in spending composition, with effect sizes less than .1% change in relative allocation per 1% enrollment decline. Instructional spending and teacher retirement decrease as a share of expenditures, while debt payments and administrative expenditures increase. Comparing raw measures of expenditures show how minor these changes are. There is null impact on instructional spending and debt payments. Teacher retirement decreases by $70,000 per 1% enrollment decrease, and administrative spending increases by

$30,000, compared to $2,640,000 and $8,130,000 in annual expenditures respectively. We conclude that there are very minor change in relative allocations of spending resulting from Save Harmless treatment. This makes our natural experiment unique, as it creates a situation where all institutional arrangements remain constant, while only student enrollment decreases. This makes educational expenditures and student-teacher ratios fundamentally equivalent for our analyses, as there is little evidence of extra money being used to fund new programs or investments.

We now proceed to direct estimation of the effect of PPE on student outcomes. The first stage models show that the Save Harmless indicator is a strong instrument for PPE, with an F-statistic of approximately 400. It is an equally strong predictor of student-teacher ratio, with an F-statistic of approximately 150. A one percent decline in student enrollment translates to a $200 increase in per-pupil expenditures and a .05 student decrease in student-teacher ratio, as reported already in Table II. Below we describe the second stage results of our 2SLS models which are reported in Table V.[2]

We estimate the effects of PPE on average math and English tests scores in grades three through eight. The results are dramatic (see Table IV). The effects in math, are highly significant with a p-value of .001.The effect sizes are meaningful at .04 standard deviations, a gain of approximately one ninth of a grade level (Bloom, Hill, Rebeck Black and Lipsey 2008). The effect of $1,000 in additional PPE on mean English scores is also 0.04 standard deviations and statistically significant with a p-value of .000. The effects of student-teacher ratios show an identical pattern, with effects of -.15 standard deviations per one student increase in student teacher ratio. Although at first glance a loss of 0.15 standard deviations per one student increase

[2] R-squared values for these models are suppressed. According to Wooldridge (2015), R-squared figures in 2SLS are often negative, and interpretations of R-squared do not lend valuable information in the context of two-stage least squares instrumental variables estimation.

appears implausibly large, this represents a change in the aggregate student teacher ratio at the district level – which has an average value of 11 students per 1 teacher. Therefore, moving from, for example 11 students per teacher to 10 students per teacher across the entire district would require a large investment, in this case about $4,000 ppe.

To compare to a commonly-used approach in the education finance literature, we estimate two-way fixed effects models of the direct effect of PPE on our dependent variables, according to equation 3. While all results are positive and statistically significant, the coefficients are much smaller, approximately between one half to one third of the size of our 2SLS estimates. The results of these analyses are presented in Table A.9. The finding that our 2SLS estimates are approximately two to three times larger supports the argument that fixed effect model estimates of the impact of PPE on student outcomes are biased downwards.

Robustness Checks

To test the robustness of our findings we first estimate models with alternate control variable strategies. Since the main threat to our model is that estimates are being driven by shifts in student demographic composition, we are particularly interested in how sensitive our estimates are to the introduction of the four demographic controls. If Save Harmless is essentially exogenous, results should be robust to different model specifications, with similar point estimates between models. We expect that standard errors could become smaller with the inclusion of covariates due to efficiency gains, but effect sizes should be reasonably consistent. The results of these analyses are included in Tables A.10 and A.11. In the PPE models, effect sizes are relatively consistent across specifications in math, and nearly perfectly consistent in English, with effect sizes changing by only approximately .007 standard deviations from the highest to lowest estimate in math (or approximately 17% of the final estimate) and shrinking by .002 standard deviations in English (or approximately 5% of the final estimate).

Trends are similar for alternative estimates of the effect of student-teacher ratios. In math the results shrink by approximately .03 standard deviations (approximately 21% of the final estimate) and shrink by .01 standard deviations in English (Approximately 8% of the final estimate). These findings are consistent with our tests in section four, where we discovered that math scores are more sensitive to changes in minority composition. These results support the conclusion that our estimates on English test scores are robust to student demographic shifts, but that our math results are potentially more sensitive.

To further probe the sensitivity of our results we estimate models with district-specific linear time trends. This modification to the model estimates a different trend slope for each district in the sample, controlling for factors that trend linearly within a district over the period of our sample. Since enrollment declined monotonically in our sample, correlated demographic shifts would also trend monotonically, making this a powerful adjustment for potential confounders[3]. The tradeoff is that this adjustment entails adding a second full vector of district dummies to the model, decreasing our degrees of freedom by approximately 650 in a model that already includes district fixed effects. In an eight-year panel, this can greatly reduce our power to detect a significant effect.

---

[3] In Table A.5 we test this correction on correlations between our instrument and demographic characteristics. We see that the instrument now achieves balance on the problematic % minority composition variable, along with % free lunch eligibility and % students with disabilities. The tradeoff is that there is now a significant relationship between % Limited English proficiency and the instrument. However, the effect is much smaller than the effect on % minority in the full models, and the relationship is positive; now limited English Proficiency students are more likely to enter the district. Since this variable has a negative impact on performance in our models with linear trends, this effect would bias our results downward.

The models with linear trends are presented in Table A.12. Effects in math, while in the same direction, are about one half the size of our main estimates, and not precisely estimated. However, results in English are of similar magnitude to our main estimates, and significant at the .01 level for PPE and at the .05 level (p= .012). These results confirm the findings of the previous section, that our results in math are sensitive, but adds greater confidence that our effects in English are robust to multiple alternative specifications. We emerge confident that educational resources do have a significant and meaningful effect, but the effects may be heterogenous between academic subjects.

Finally, we test the sensitivity of our estimates to between-district differences in enrollment trends. One endogeneity concern that we discussed in section 4 was that the effects might be driven by the negative impacts of enrollment increase. To ensure this is not the case, we restrict our sample to only observations in which districts experienced an enrollment decrease. This ensures that no aspect of the result is being driven by comparisons to districts with enrollment increase. The results of these tests are included in Table A.13. The results in this subsample are fundamentally equivalent in effect size and significance to those in the main models. This demonstrates that the effects detected are entirely driven by districts maintaining higher funding levels through the Save Harmless provision, indicating that we have effectively operationalized our instrument.

6. Discussion

The findings in this study show clear and compelling evidence that educational resources improve student learning, providing a new contribution to an old debate over the relationship between educational resources and student performance. This study leverages a natural experiment in the context of NYS school finance mechanics to demonstrate the relationship.

While more research is needed to resolve this debate across different contexts and with different empirical strategies, our results, coupled with other findings from quasi-experimental studies, should suggest that allowing districts to maintain high levels of funding even during periods of enrollment loss benefits student performance.

Efforts to deliver additional resources to low-performing schools have been a long-standing endeavor among educational reformers. After succeeding in equalizing raw measures of expenditures per pupil, advocates advanced the further argument that under-performing school districts actually needed more money than wealthy districts to achieve the same level of performance (Bifulco 2005; Duncombe and Yinger 2000; Lafortune, Rothstein and Schanzenback 2016; Yinger 2004). While such performance-based educational finance reforms have flourished in recent years, the field still lacked credible evidence that increases in spending could meaningfully change educational performance and help close these gaps. With inconclusive findings in the economics of education literature, critics of school spending have been able to argue that such reforms are futile and misguided.

The findings of this study provide evidence that undermines these critiques. They show that even small increases in PPE, on the order of $1,000 per pupil, or less than five percent of current average education spending in New York, can increase educational performance by approximately 0.04 standard deviations. These effects are twice as large as recent estimates provided by Lafortune et. al (2016), which could be explained by the differing contexts or the fact that all increased resources went into reduced class size in this case[4]. As achievement gaps

[4] Lafortune et al. (2016) use ten years of spending as their independent variable, and so we compare our one-year effect size to their estimates divided by ten. Other recent high quality studies are more difficult to compare our effect sizes to given that they either did not focus on test score outcomes or provided test score outcomes in terms of proficiency rates.

between privileged and disadvantaged students remain a persistent source of inequality in society (Clotfelter, Ladd and Vigdor 2009; Fryer and Levitt 2004; Fryer and Levitt 2006; Hanushek 1997; Reardon 2011), this study provides a persuasive case that greater investment in education could help close these gaps. Coupled with other credible results from quasi-experimental analyses, (Jackson, Johnson and Persico 2016; Papke 2005; Guryan, 2001) this study builds onto a growing body of evidence that educational resources contribute to improved student outcomes. Nonetheless, the results suggest that the effects of spending are not large enough to completely solve the problem of academic achievement gaps. Closing these gaps, which currently measure approximately 1 standard deviation, would require increasing PPE by around 115%, which would likely be practically and politically infeasible even if only implemented in targeted districts. Nonetheless, these findings caution against spending cuts, and support targeted expenditure increases as a policy instrument to improve academic performance.

These results also have policy implications for the effectiveness or desirability of Save Harmless policies. Save Harmless policies are controversial, because critics say they divert resources from real students with real needs to support students that don’t exist (Chira 1983; Cunningham 2014; Levine 1976). However, our findings show that while these policies may contribute to equity losses in theory, in practice they can have broad and equally-distributed impacts on student achievement. In the context of our study, nearly every district in NYS benefitted from additional financial aid under the Save Harmless policy, and experienced notable improvements in student achievement. In a state like New York, which is threatened by serious demographic changes which could undermine the stability of school districts state wide, Save Harmless policies helped schools weather the disruptions. These results may not generalize to other states, where there is more pronounced variation in areas with and without enrollment decline, which could lead to unevenly distributed benefits. But in states where the policy's impacts are evenly distributed, Save Harmless policies can mitigate adverse demographic trends through resource stability.

This study is subject to some limitations. The quasi-experimental methods employed in this study are subject to rigorous assumptions, the validity of which we may not be able to perfectly defend given the limitations of our data, and the unique conditions of the natural experiment upon which our analysis is based. We probe the sensitivity of our estimates to violations of these assumptions, and find that they have limited impact on our estimates, though our estimates of effects on English scores may be more credible than those on math scores. As in all cases of empirical analysis, the results of this study cannot provide the definitive statement on the long-standing question of the relationship between educational resources and student performance. But they do construct a narrative of how, in this case, state aid maintenance over time resulted in smaller class sizes and marked educational improvement.

We offer more current estimates of the causal impact of PPE than the studies we have cited, using data as recent as the 2014-15 academic year. We also offer the first estimates to our knowledge of the impact of PPE in New York State following their 2007 educational finance reform, which allows our research to inform discussions of school finance reforms at both the state and national level. Finally, this study provides the first examination of the effects of Save Harmless policies on student outcomes, which makes a salient contribution to the literature on school finance systems. We have investigated a novel approach for isolating variation in educational aid resulting from a particular school finance reform provision; future research could investigate other state-level finance formulas for similar idiosyncratic provisions to shed light on this debate across varying fiscal and institutional contexts. In recent years, 22 percent of total state and local spending have gone to funding elementary and secondary education (US Census, 2014). Continuous and rigorous analysis of the benefits that accrue to students can inform how decision-makers allocate resources to and within the education sector during times of fiscal scarcity.

References

Abbott, Doug, Patrick Hodgens, and Kevin Wenzel, "Memorandum on New York State Education Aid Formula Reform," (Center for Policy Research, Syracuse University, 2013).

Angrist, Joshua D., and Victor Lavy, "Using Maimonides' rule to estimat the effect of class size on scholastic achievement," Quarterly Journal of Economics, 114 (1999), 533-575.

Bifulco, Robert, "District-level black-white funding disparities in the United States, 1987-2002," Education Finance and Policy, 31 (2005), 172-194.

Bloom, Howard S., Carolyn J. Hill, Alison Rebeck Black, and Mark W. Lipsey, "Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions," Journal of Research on Educational Effectiveness, 1 (2008), 289-328.

Brummet, Quentin. "The effect of school closings on student achievement." Journal of Public Economics, 119, 108-124.

Card, David, and A. Abigail Payne, "School finance reform, the distribution of school spending, and the distribution of student test scores," Journal of Public Economics, 83 (2002), 49-82.

Chira, Susan, ""Save Harmless" School Formula Survives Anew," in *The New York Times*, (New York, NY, 1983).

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "The academic achievement gap in grades 3 to 8," The Review of Economics and Statistics, 91 (2009), 398-419.

Cunningham, Deborah H., "State aid to school districts in New York State: An overview based on the laws of 2014," (New York State Association of School Business Officials, 2014).

Duncan, Greg G., and Richard J Murnane, eds., *Whither Opportunity?: Rising Inequality, Schools and Children's Life Chances.* (New York, NY: Russell Sage Foundation, 2011).

Duncombe, William, and John Yinger, "Financing higher student performance standards: the case of New York State," Economics of Education Review, 19 (2000), 363-386.

Empire Center, "Population is dropping faster in upstate New York counties," (2016).

Fryer, Roland G., and Steven D. Levitt, "Understanding the black-white test score gap in the first two years of school," The Review of Economics and Statistics, 86 (2004), 447-464.

---, "The black-white test score gap through third grade," American Law and Economics Review, 8 (2006), 249-281.

Greenwald, Rob, Larry V. Hedges, and Laine Richard D., "The effect of school resources on student achievement," Review of Educational Research, 66 (1996), 361-396.

Guryan, Jonathan, "Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts," NBER Working Paper: No. 8269, (2001).

Hanushek, Eric A., "A jaundiced view of "adequacy" in school finance reform," Educational Policy, 8 (1994), 460-469.

---, "Assessing the effects of school resources on student performance: An update," Educational Evaluation and Policy Analysis, 19 (1997), 141-164.

Hanushek, Eric A., and Steven G. Rivkin, "School quality and the black-white achievement gap," (National Bureau of Economic Research: Working Paper No. 12651, 2006).

Heckman, James J., and Paul A. LaFontaine, "The American high school graduation rate: Trends and levels," The Review of Economics and Statistics, 92 (2010), 244-262.

Hoxby, Caroline M., "The effects of class size on student achievement: New evidence from population variation," Quarterly Journal of Economics, 115 (2000), 1239-1285.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico, "The effects of school spending on educational and economic outcomes: Evidence from school finance reforms," The Quarterly Journal of Economics, 131 (2016), 157-218.

Kotkin, Joel, "The states gaining and losing the most migrants -- and money," in *Forbes*, (2016).

Krueger, Alan B., "Experimental estimates of educational production functions," Quarterly Journal of Economics, 114 (1999), 497-532.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenback, "School finance reform and the distribution of student achievement," NBER Working Paper: No. 22011, (2016).

Levine, Harold, "The Harm of "Save Harmless"," in *The New York Times*, (New York, NY, 1976).

NYS Department of Labor, "Employment in New York State," (2011).

Papke, Leslie, "The effects of spending on test pass rates: Evidence from Michigan," Journal of Public Economics, 89 (2005), 729-1154.

Reardon, Sean F., "The widening academic achievement gap between the rich and the poor: New evidence and possible explanations," in *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*, Greg G. Duncan, and Richard J Murnane, eds. (New York, NY: Russell Sage Foundation, 2011).

Rockoff, Jonah E, and Benjamin B. Lockwood. "Stuck in the middle: Impacts of grade configuration in public schools." Journal of Public Economics, 94 (11-12), 1051-1061.

Schwartz, Amy Ellen, Stiefel, Leanna, and Sarah A. Cordes. "Moving matters: The causal effect of moving schools on student performance." Education Finance and Policy, 12(4), 419-446.

Verstegen, Deborah A., and Richard A. King, "The relationship between school spending and student achievement: A review of 35 years of production function research," Journal of Education Finance, 24 (1998), 243-262.

Wing, Paul, "Population trends in New York State: New Yorkers at the Millennium," in *The Public Policy Institute*, (2003).

Wooldridge, Jeff, *Econometric Analysis of Cross Section and Panel Data* 2 ed. (Cambridge, Massachusetts: The MIT Press, 2010).

---, *Introductory Econometrics: A Modern Approach* 6 ed. (Cengage Learning, 2015).

Yinger, John, ed., *Helping Children Left Behind: State Aid and the Pursuit of Educational Equity* (Cambridge, MA: MIT University Press, 2004).

Tables and Figures

Figure I: Patterns in Changes in Student Enrollment and Total Expenditures Across NYS

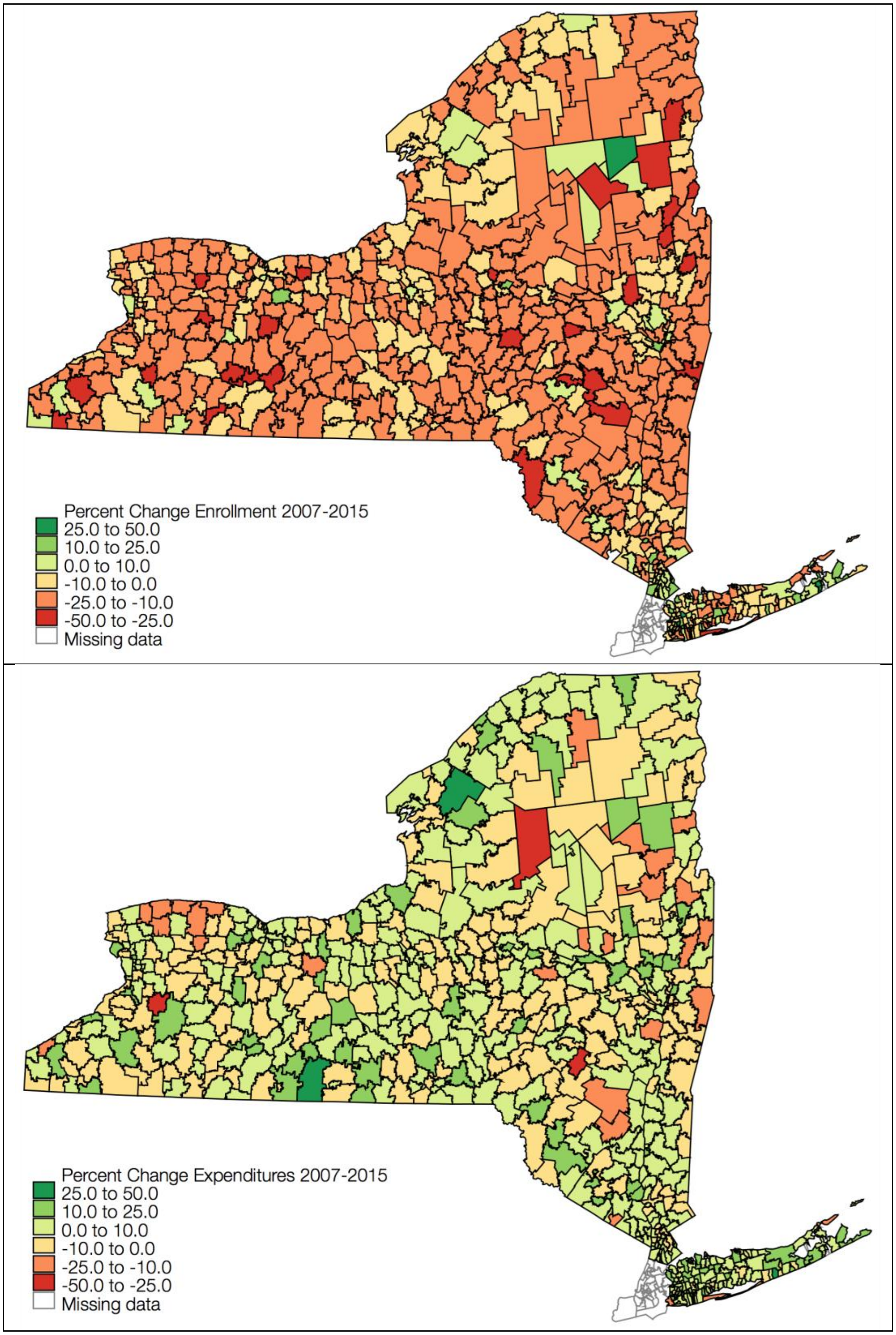
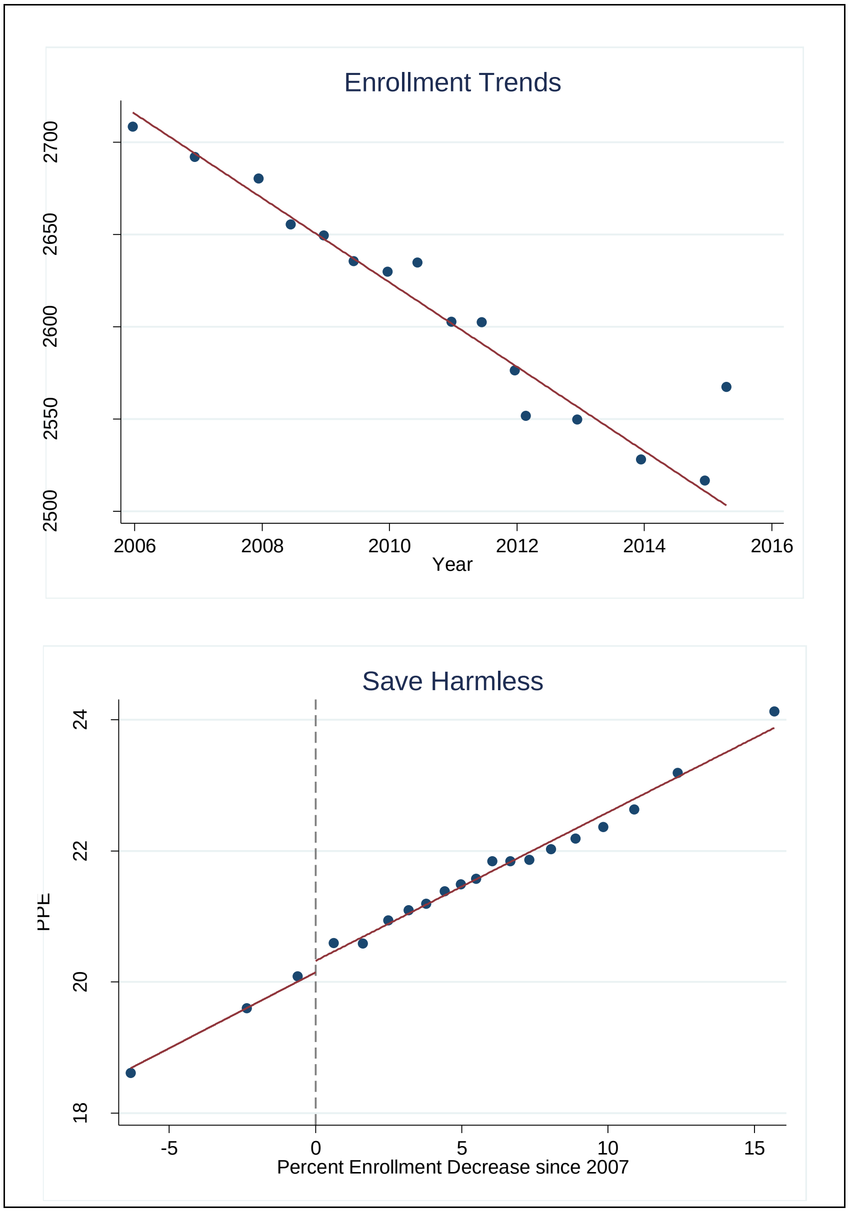


Figure II: Declining Enrollment and the Relationship between Enrollment Change and PPE



Note. These figures are binned scatter plots absorbing district fixed effects with linear trend lines added. The first graph plots average district student enrollment across New York between 2008 and 2016; the second graph plots per-pupil expenditures (PPE) by the percent enrollment decline since the 2006-2007 school year.

Table I: Descriptive Statistics					
Variable	Obs	Mean	Std. Dev.	Min	Max
Math Scores	5116	0.00	1.00	-3.58	3.27
English Scores	5116	0.00	1.00	-3.46	3.37
Save Harmless	5116	6.59	6.41	0.00	36.28
PPE	5116	21.70	4.841229	12.64746	70.8227
Student-Teacher Ratio	5116	11.39	1.66	4.21	25.00
State Aid PP	5116	8.63	4.08	0.92	23.17
Instructional Expenditures	5116	30299.09	39549.09	1655.81	461850.40
% Instructional	5116	0.72	0.05	0.29	0.90
Teacher Retirement	5116	2224.16	2948.67	64.21	35051.26
Teacher Retirement Ratio	5116	0.04	0.01	0.01	0.08
Debt Expenditures	5116	3694.67	5961.59	0.00	186242.40
Debt Ratio	5116	0.08	0.04	0.00	0.67
Administrative Expenditures	5116	7466.61	9640.99	457.94	123104.60
Administrative Ratio	5116	0.14	0.02	0.05	0.29
% Free Lunch Eligibility	5116	25.88	15.94	0.00	95.00
% Minority	5116	17.00	20.34	0.00	100.00
% LEP	5116	1.99	4.13	0.00	33.00
% Students with Disabilities	5116	15.58	4.71	3.21	52.67

Table II: First Stage Models			
	(1)	(2)	(3)
VARIABLES	PPE	State Aid PP	Student Teacher Ratio
Save Harmless	0.2080**	0.0909**	-0.0578**
	(0.0104)	(0.0055)	(0.0047)
Enrollment	-0.0008**	0.0001	0.0004**
	(0.0002)	(0.0001)	(0.0002)
Administrative Ratio	-21.5481**	2.1284	4.8399**
	(5.5210)	(1.6227)	(1.2640)
Debt Ratio	26.0501**	8.4164**	1.6208**
	(5.8589)	(1.1367)	(0.4779)
% Free Lunch Eligibility	-0.0091+	-0.0015	0.0073**
	(0.0047)	(0.0024)	(0.0023)
% Minority	0.0215**	0.0003	-0.0007
	(0.0067)	(0.0030)	(0.0034)
% LEP	-0.0388	-0.0290+	0.0039
	(0.0349)	(0.0151)	(0.0453)
% Students with Disabilities	0.0078	0.0001	-0.0070**
	(0.0065)	(0.0029)	(0.0025)
Constant	22.9681**	6.9857**	9.2929**
	(0.8527)	(0.4789)	(0.4858)
Observations	5,116	5,116	5,116
R-squared	0.6209	0.5015	0.1956
F-Statistic (Save Harmless)	398.91	268.61	152.62
Year FE	x	x	x
District FE	x	x	x
Number of District	649	649	649

Robust t-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1

Table III: Effects on Educational Expenditure Allocation				
	(1)	(2)	(3)	(4)
VARIABLES	Instructional Expenditures	Instructional Ratio	Retirement	Retirement Ratio
Save Harmless	11.4098	-0.0005**	-71.5998**	-0.0002**
	(13.1881)	(0.0001)	(10.6621)	(0.0000)
Observations	5,116	5,116	5,116	5,116
R-squared	0.2209	0.4468	0.4534	0.9277
Year FE	x	x	x	
District FE	x	x	x	
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Table IV: Effects on Educational Expenditure Allocation				
	(1)	(2)	(3)	(4)
VARIABLES	Debt Expenditures	Debt Ratio	Administrative Expenditures	Administrative Ratio
Save Harmless	2.4566	0.0007**	29.7480**	0.0003**
	(9.9594)	(0.0002)	(5.2351)	(0.0001)
Observations	5,116	5,116	5,116	5,116
R-squared	0.0505	0.1590	0.1737	0.3212
Year FE	x	x	x	
District FE	x	x	x	
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Table V: 2SLS Estimates of the Effect of PPE And Student Teacher Ratio on Test Scores				
	(1)	(2)	(3)	(4)
VARIABLES	Math	Math	English	English
PPE	0.0411**		0.0399**	
	(0.0129)		(0.0105)	
Student Teacher Ratio		-0.1478**		-0.1435**
		(0.0460)		(0.0388)
Enrollment	0.0001	0.0001+	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Administrative Ratio	0.7317	0.5612	0.9966	0.8310
	(0.8316)	(0.7733)	(0.7062)	(0.6465)
Debt Ratio	-0.8065+	0.5039+	-0.9614*	0.3112
	(0.4698)	(0.2762)	(0.4198)	(0.2424)
	-		-	
% Free Lunch Eligibility	0.0055**	-0.0048**	0.0042**	-0.0035**
	(0.0013)	(0.0013)	(0.0012)	(0.0012)
	-			
% Minority	0.0072**	-0.0064*	-0.0019	-0.0012
	(0.0027)	(0.0026)	(0.0017)	(0.0017)
% LEP	-0.0262*	-0.0272*	-0.0014	-0.0024
	(0.0116)	(0.0114)	(0.0101)	(0.0120)
% Students with Disabilities	-0.0022	-0.0029+	-0.0029+	-0.0036*
	(0.0017)	(0.0017)	(0.0016)	(0.0016)
			-	
Constant	-0.8466*	1.4710**	0.9024**	1.3483**
	(0.4309)	(0.4115)	(0.3360)	(0.3467)
Observations	5,116	5,116	5,116	5,116
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649
Robust z-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Appendix

Data Construction Supplementary Information

Percent free lunch eligibility included one value greater than 100, we replace this value with the mean of the prior year and the following year observations.

Approximately 200 observations for percent students with disabilities were systematically suppressed in years 2012-13 and 2013-14. To preserve these variables, we used Stata 14.2 SE's impute command, which is a regression based imputation that uses observed information to fill in missing values. The imputation for the students with disabilities variable was calculated using the following variables: number of students in each grade 3-8, percent minority, percent free lunch eligibility, standardized test scores for grades 3-8 math and English, standardized math and English performance indices, gross special education expenditures (unadjusted) and district expenditures per pupil (unadjusted). A comparison of the original vs. the imputed variables is provided in Appendix Table 1.

The test score dependent variables were missing between 50 and 100 69 and   observations each without complete overlap between missing observations. For this reason, we restrict our analysis sample to only those observations for which both variables are not missing. This leads to a total of approximately 100 missing observations for each variable, out of approximately 5,200 possible observations.

Appendix Table 1: Comparison of Original and Imputed Value

Variable	Obs	Mean	Std. Dev.	Min	Max
% SWD (Original)	4905	16	4.79	3.21	52.67
% SWD (Imputed)	5116	16	4.71	3.21	52.67

Appendix Table 2: Conditional Independence				
	(1)	(2)	(3)	(4)
VARIABLES	%Free Lunch	% Minority	% LEP	%SWD
Save Harmless	0.0045	-0.1147**	-0.0197**	-0.0002
	(0.0304)	(0.0231)	(0.0068)	(0.0002)
Observations	5,116	5,116	5,116	5,116
R-squared	0.4503	0.2388	0.0478	0.0633
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Appendix Table 3: Conditional Independence (With Controls)				
	(1)	(2)	(3)	(4)
VARIABLES	%Free Lunch	% Minority	% LEP	%SWD
Save Harmless	0.0304	-0.1563**	-0.0069	-0.0003
	(0.0336)	(0.0232)	(0.0055)	(0.0003)
Observations	5,116	5,116	5,116	5,116
R-squared	0.4566	0.2590	0.0764	0.0664
Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Appendix Table 4: Conditional Independence (With Controls & Linear Trends)				
	(1)	(2)	(3)	(4)
VARIABLES	%Free Lunch	% Minority	% LEP	%SWD
Save Harmless	-0.0401	-0.0116	0.0075*	0.0003
	(0.0579)	(0.0281)	(0.0037)	(0.0003)
Observations	5,116	5,116	5,116	5,116
R-squared	0.6007	0.5781	0.6242	0.3144
Covariates	x	x	x	x
Year FE	x	x	x	x

District FE	x	x	x	x
District Specific Linear Trends	x	x	x	x
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Appendix Table 5: Conditional Independence (Predicted PPE)

VARIABLES	(1) %Free Lunch	(2) % Minority	(3) % LEP	(4) %SWD
PPE	0.1466	- 0.7638**	-0.0330	-0.0014
	(0.1625)	(0.1248)	(0.0266)	(0.0012)
Observations	5,116	5,116	5,116	5,116
Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Appendix Table 6: Conditional Independence (Predicted Student Teacher Ratio)

VARIABLES	(1) %Free Lunch	(2) % Minority	(3) % LEP	(4) %SWD
Student Teacher Ratio	-0.5284	2.7065**	0.1186	0.0051
	(0.5883)	(0.4572)	(0.0962)	(0.0044)
Observations	5,116	5,116	5,116	5,116
Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649
Robust t-statistics in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Appendix Table 7: Monotonicity (PPE)

VARIABLES	(1) PPE (Urban High Need)	(2) PPE (Rural High Need)	(3) PPE (Medium Need)	(4) PPE (Low Need)
Save Harmless	0.179**	0.233**	0.210**	0.296**
	(0.0397)	(0.0167)	(0.0143)	(0.0405)

Observations	375	1,216	2,630	894
R-squared	0.459	0.162	0.175	0.207
Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	53	172	362	125

Robust t-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1

Appendix Table 8: Monotonicity (Student Teacher Ratio)

VARIABLES	(1) PPE (Urban High Need)	(2) PPE (Rural High Need)	(3) PPE (Medium Need)	(4) PPE (Low Need)
Save Harmless	-0.0737*	-0.0614**	-0.0653**	-0.0581**
	(0.0298)	(0.00954)	(0.00642)	(0.0105)
Observations	375	1,216	2,630	894
R-squared	0.459	0.162	0.175	0.207
Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	53	172	362	125

Robust t-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1

Appendix Table 9: Two-Way Fixed Effects Estimates of the Effects of Educational Resources on Student Achievement

VARIABLES	(1) Math	(2) Math	(3) English	(4) English
PPE	0.0170**		0.0138**	
	(0.0050)		(0.0043)	
Student Teacher Ratio		-0.0707**		-0.0467**
		(0.0127)		(0.0093)
Observations	5,116	5,116	5,116	5,116
R-squared	0.0254	0.0382	0.0126	0.0181
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649

Robust t-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1

Appendix Table 10: Alternate Control Variable Strategies (PPE)						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Math	Math	Math	English	English	English
PPE	0.0463**	0.0480**	0.0410**	0.0400**	0.0416**	0.0398**
	(0.0126)	(0.0132)	(0.0129)	(0.0102)	(0.0107)	(0.0105)
Enrollment	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Administrative Ratio		0.8234	0.7232		0.9602	0.9887
		(0.8453)	(0.8315)		(0.7039)	(0.7060)
Debt Ratio		-0.9535+	-0.8069+		-1.0011*	-0.9602*
		(0.5001)	(0.4697)		(0.4305)	(0.4198)
% Free Lunch Eligibility			-0.0055**			-0.0042**
			(0.0013)			(0.0012)
% Minority			-0.0071**			-0.0019
			(0.0027)			(0.0017)
% LEP			-0.0264*			-0.0015
			(0.0116)			(0.0101)
% Students with Disabilities			-0.0022			-0.0029+
			(0.0017)			(0.0016)
Constant	-1.2113**	-1.3123**	-0.8387+	-0.9814**	-1.0988**	-0.8940**
	(0.3912)	(0.4369)	(0.4314)	(0.3012)	(0.3387)	(0.3362)
Observations	5,116	5,116	5,116	5,116	5,116	5,116
Year FE	x	x	x	x	x	x
District FE	x	x	x	x	x	x
Number of District	649	649	649	649	649	649
Robust z-statistics in parentheses						
** p<0.01, * p<0.05, + p<0.1						

Appendix Table 11: Alternate Control Variable Strategies (Student Teacher Ratio)						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Math	Math	Math	English	English	English
Student Teacher Ratio	-0.1787**	-0.1708**	-0.1474**	-0.1542**	-0.1481**	-0.1432**
	(0.0486)	(0.0469)	(0.0460)	(0.0403)	(0.0390)	(0.0388)
Enrollment	0.0002*	0.0001*	0.0001+	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Administrative Ratio		0.6323	0.5531		0.7944	0.8234
		(0.7758)	(0.7733)		(0.6384)	(0.6465)
Debt Ratio		0.5728*	0.5001+		0.3228	0.3095
		(0.2770)	(0.2763)		(0.2418)	(0.2426)
% Free Lunch Eligibility			-0.0049**			-0.0036**
			(0.0013)			(0.0012)
% Minority			-0.0064*			-0.0012
			(0.0026)			(0.0017)
% LEP			-0.0274*			-0.0025

			(0.0114)			(0.0120)
% Students with Disabilities			-0.0029+			-0.0036*
			(0.0017)			(0.0016)
Constant	1.5909**	1.3943**	1.4728**	1.4370**	1.2489**	1.3515**
	(0.4434)	(0.4182)	(0.4118)	(0.3633)	(0.3501)	(0.3469)
Observations	5,116	5,116	5,116	5,116	5,116	5,116
Year FE	x	x	x	x	x	x
District FE	x	x	x	x	x	x
Number of District	649	649	649	649	649	649

Robust z-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1

Appendix Table 12: District Specific Linear Trends

VARIABLES	(1) Math	(2) Math	(3) English	(4) English
PPE	0.0127		0.0298**	
	(0.0135)		(0.0115)	
Student Teacher Ratio		-0.0633		-0.1487*
		(0.0666)		(0.0593)
Observations	5,116	5,116	5,116	5,116
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	649	649	649	649

Robust t-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1

Appendix Table 13: Restricted Sample (Decline Only)

VARIABLES	(1) Math	(2) Math	(3) English	(4) English
PPE	0.0416**		0.0384**	
	(0.0138)		(0.0118)	
Student Teacher Ratio		-0.1489**		-0.1374**
		(0.0490)		(0.0427)
Observations	4,143	4,143	4,143	4,143
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	610	610	610	610

Robust t-statistics in parentheses
** p<0.01, * p<0.05, + p<0.1