

Educational Expenditures 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 expenditures 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 controversy results from a shortage of experimental or quasi-experimental analyses to identify the relationship. This study leverages exogenous variation in per-pupil expenditures from a specific provision of the state aid formula in New York State to identify the impact of school spending on student outcomes. We uncover performance gains of between approximately .05 and .07 standard deviations in both math and English corresponding to \$1,000 in additional per-pupil spending, when expenditures are targeted to economically disadvantaged and struggling students. This study strengthens the case that school resources matter, and that targeted financial investments can help close educational achievement gaps. I22, C36, D00

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 widespread

improvements in disadvantaged public school districts, and researchers still struggle to reliably establish the link between spending and performance. This leads to disagreement among experts over whether educational expenditures have any causal positive impact on student performance. (Greenwald, Hedges and D. 1996; Hanushek 1997).

Hanushek (1997) and Greenwald, Hedges and D. (1996) each performed meta-analyses of studies connecting educational expenditures to student performance, summarizing a literature riddled with contradictory results and statistically insignificant findings. Even after analyzing the same set of studies, they arrived at differing conclusions. While Greenwald and colleagues argued that the literature indicates an overall positive relationship between spending and performance, Hanushek offers the contention that the research is inconclusive at best. One problem with many of these early studies is that they rely on endogenous estimates of educational expenditure effects. However, a few recent studies have exploited natural experiments to derive more credible estimates of the impact of spending on performance. Jackson, Johnson and Persico (2016), Papke (2005); Papke (2008) and Chaudhary (2009) use exogenous variation in state educational aid to identify positive impacts of educational resources on student outcomes. However, since these are the only quasi-experimental treatments of this subject to our knowledge, more evidence is needed to support a definitive conclusion that generalizes across states and fiscal contexts.

This paper contributes to the literature by employing quasi-experimental methods to investigate the relationship between educational expenditures and student performance in New York State (NYS) school districts between the 2007-08 through 2014-15 academic years. It relies on exogenous variation in district level state aid to identify impacts of per-pupil expenditures on student performance. 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 declining enrollment trends to be consistent across New York School districts. We demonstrate that enrollment change was uncorrelated with 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 per-pupil expenditures (PPE).

Leveraging this variation through instrumental variable estimation, we find positive effects of expenditures on elementary and middle school academic proficiency that apply to the economically disadvantaged and struggling student populations. We probe the validity of these inferences, specifically the validity of our exclusion restriction assumptions, through a comprehensive set of conditional exogeneity tests and robustness checks that demonstrate that Save Harmless treatment occurs quasi-randomly conditional on district and year fixed effects and district enrollment, producing plausibly exogenous variation in per-pupil expenditures.

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 per-pupil expenditures in New York State following their 2007 school finance reform. Since this 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 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 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. Results become more significant at higher levels of aggregation. 17% of school level studies show a positive and statistically significant coefficient, while 28% of district level studies and 64% of state level studies do. He concludes that there is no consistent evidence that expenditures have a positive effect on educational achievement, and thus policymakers should have little confidence that increasing spending will improve outcomes in low-performing schools.

Other researchers are more confident in the ability of expenditures to influence student outcomes. 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. 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.

This study will contribute to the literature by developing a novel approach for identifying effects of increased per-pupil expenditures. Specifically, by leveraging exogenous 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 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 essentially random variation in per-pupil expenditures. We defend this argument through a rigorous set of falsification tests and descriptive analyses.

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, which spans the 2007-08 to 2014-15 academic years, the mean year-by-year percentage change in enrollment was -1.35%, with 71% of observations in

our sample showing declining enrollment. The mean percentage change in observations with declining enrollment was -2.76%. Out of 652 districts in our analysis sample, only five experienced zero years of declining enrollment during the period of our study. To assess these enrollment trends and their impact on district level financial and demographic characteristics we present descriptive graphics in Figure I.

The graphics in Figure 1 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 a that enrollment declined year over 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 enrollment change 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 marked 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. However, the plots in Figure II demonstrate that this wasn't the case. These graphs plot the relationship between percentage enrollment change and 4 measures of district demographic composition: percent free lunch eligibility, percent minority status, percent limited English proficiency and percent students with disabilities. All graphs show relatively flat relationships with enrollment change, providing no evidence which would elicit concerns about simultaneous district-level demographic change.

Ultimately these graphical analyses demonstrate consistent enrollment trends across NYS district types, and a remarkably consistent relationship between within-district yearly enrollment change and within-district changes in per-pupil expenditures, bolstering the case that the Save Harmless provision, and state aid formula more generally, have substantive consequences for school funding levels. They also demonstrate that there are no clear demographic consequences of enrollment change that could confound the relationship between increased PPE and student performance. In the methods section, we will test more rigorously whether within-district enrollment declines cause changes to the student population other than through expenditure increases, and then fully account for any such compositional shifts.

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 year prior. 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 revenues. It is calculated by first subtracting enrollment in the prior period from enrollment in the current period, and dividing by enrollment in the prior period. If there is a decline in enrollment from the prior period, this variable will have a negative value. We therefore multiply the measure by negative one, so that coefficients will be positive if there is a positive relationship between yearly declining enrollment and expenditures or student outcomes.

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 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. The student outcomes are average mathematics and English standardized tests scores for grades three through eight, and mathematics and English performance indices. Performance indices are measures of students in grades three through eight achieving proficiency on annual standardized tests. Performance indices are the key accountability measure used to assess yearly progress in NYS schools. Measured out of 200, they are a ratio of students achieving proficiency according to six proficiency levels, which penalizes districts for having students in low-performing categories without evidence of progress. We also include test score measures pertaining to only economically disadvantaged students. This is possible because NYS school districts report performance measures pertaining not only to the general population, but to student subgroups as well. The test score and performance index scales changed during 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 increased per-pupil expenditures on student performance, assigning the Save Harmless policy as an instrument for expenditures. These primary methods, along with a comprehensive set of robustness tests and conditional exogeneity tests, are described in this section.

4.1 Estimating the Effect of Enrollment Declines on Institutional Outcomes

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:

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

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 a vector of coefficients on a vector of enrollment terms (DCAADM and Quadratic 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 increased per-pupil educational expenditures (PPE) 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, 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 impact 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.

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

(Equation 3)

In this equation, *PPE* is our measure of expenditures per pupil 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 a vector of enrollment characteristics, 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 established earlier in this section and further defended in the following tests of the validity of our instrument, then these predictions of expenditures per pupil are un-confounded by unobservable impacts of *Save Harmless*. We can

use them to identify the causal impact of PPE on performance in a second stage equation of the following form:

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

(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 in \widehat{PPE} , our predicted values of PPE. All other terms are identical to the first stage model. If our assumptions hold, then δ_1 will reflect the true impact of PPE 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 test 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. When evaluating a randomized control trial, it is standard to check for the exogeneity of the experimental treatment by performing a randomization test to assess whether the treatment is correlated with sample covariates. We perform a similar test by regressing four demographic characteristics on our measure of Save Harmless treatment. In models with district and year fixed effects, the instrument has no significant effect on any measure of student characteristics (Table A.2). We also estimate these models with the full set of enrollment terms and district-level covariates as specified in equation 1 (Table A.3). Once more, all coefficients on Save Harmless are statistically insignificant. This implies that year-to-year enrollment declines within districts in our sample do not cause any significant shifts in the composition of students. While it is therefore not strictly necessary to control for student characteristics in our model, we include them in our 2SLS models to attain resulting efficiency gains.

Exclusion Restriction

For *Save Harmless* to serve as an instrument, it must only impact student outcomes through its effect on PPE. Because Save Harmless treatment is triggered by declining enrollment which may be correlated with unobservable characteristics of school districts that also predict student performance, we must carefully account for alternative mechanisms. We address this in the following ways. First, we control for enrollment and polynomial enrollment terms to isolate the impact of the Save Harmless provision that is not related to the effects of current district student enrollment. We also include district fixed effects, which control for all time-invariant observed and unobserved district characteristics, and year fixed effects which control for all time trends across the years in our sample. We believe that this is a powerful approach for eliminating alternate mechanisms by which Save Harmless treatment may affect student performance

measures, and in using this approach we can identify plausibly exogenous variation in school spending.

Skeptics may note that districts subject to the Save Harmless policy, who experience declining enrollment, may also experience simultaneous demographic changes which would affect student performance levels. Exploratory analyses in the prior section (Tables A.2 & A.3) show that we should not be overly concerned with unobservable student confounders. Nonetheless, our model controls for this alternate mechanism with four key demographic indicators: percent of students free lunch eligible, percent of students minority, percent of students limited English proficiency, and percent of students with disabilities.

We can test (albeit, imperfectly) the validity of the exclusion restriction directly using a simple test proposed by Wooldridge (2010). If an independent variable is strictly exogenous conditional on a fixed effect, it should not be correlated with confounding characteristics that carry between periods. Therefore, even if a variable has enduring impacts that carry forward in time, it would not be possible for future observations of the variable to have an impact on prior periods. Under strict exogeneity, the parameter β_1 in the following equation should be zero:

$$y_{dt} = \beta_1 \text{Save Harmless}_{dt+1} + \beta_2 \text{Save Harmless}_{dt} + \beta_3 X_{dt} + \theta_d + \tau_t + \varepsilon_{dt}$$

(Equation 2)

In this equation the vector X_{dt} includes current student enrollment. We estimate equation 2 for each of our six student outcomes in all twelve model variations presented in the results section. We also adapt this test to the assumptions of our model by including lead covariates and lead enrollment terms. The results of these analyses are included in Tables A.4, A.5 and A.6. All

results are statistically insignificant with absolute values t-statistics less than 1 in 10 out of 12 models. This contributes persuasive evidence for our conditional exogeneity argument.

We can also adapt this test to a 2SLS approach. We use a one-year lead value of *Save Harmless* as an instrument for one-year lead PPE, and include one-year leads of our enrollment terms and covariates for an alternative first stage, and then we estimate the same second stage approach as specified in section 4.2 – this time with the alternative predicted PPE values. (We also include present year covariates, so our models are comparable to those in equations 3 and 4). These models constitute a placebo test, where we demonstrate that future expenditures, which cannot temporally have an impact on student outcomes, are not correlated with our outcomes. The results of these analyses are included in Tables A.7, A.8 and A.9. All twelve models show null results with absolute value t-statistics less than 1 in 10 out of 12 models, providing additional evidence for the validity of our exclusion restriction.

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 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. We also estimate an alternative model with only observations in which the district experienced either an enrollment increase or no

change in enrollment, and was therefore untreated by the Save Harmless policy. The results are included in Table A.10. The results show a positive relationship of enrollment decline with PPE 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 per-pupil expenditures and student-teacher ratios in NY school districts. In the second step, we employ a two-stage least squares (2SLS) approach to identify the direct impacts of PPE 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 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, and columns 2 and 4 contain additional control measures of quadratic current student enrollment. Columns 1 and 2 show 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 \$110 dollars per pupil per 1% decline in enrollment. This finding carries over to measures of student-teacher ratios. Columns 3 and 4 show a statistically significant decline in student teacher ratios of approximately .013 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.

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 F-statistics greater than 140 in all models (See Table III). A one percent decline in student enrollment translates to a \$110 increase in per-pupil expenditures, as reported already in Table 5. Below we present the second stage results of our 2SLS models which are reported in tables 4, 5 and 6.¹

First, we estimate the effects of PPE on average math and English tests scores in grades three through eight. The results are moderate (see Table IV). The effects in math, while statistically significant at the .01 level, are relatively small 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 0.02 standard deviations and statistically insignificant. However, when we examine results that only represent economically disadvantaged students, we observe a different phenomenon. Results pertaining to economically disadvantaged students are reported in Table IV. Results are now statistically significant at the .05 level in both math and English and twice as large, at .07 standard deviations, or one fifth of a grade level. This indicates that school spending has a larger effect on the academic outcomes of disadvantaged students. We confirm this with our analysis of the effect of PPE on math and English performance indices. As we described in the data section, performance indices penalize districts for the number of students in the lowest proficiency levels on standardized tests who are not making progress. For this reason, they are a good measure of how the lower-performing students in a district are doing. The results of these analyses are presented in Table VI. Effects on

¹ R-squared values for these models are suppressed. According to Wooldridge, Jeff, *Introductory Econometrics: A Modern Approach* 6 ed. (Cengage Learning, 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.

math and English are both significant at the .01 level, with effect sizes between .049 and .066 standard deviations. The magnitudes of these student achievement effects are large enough to justify increasing expenditures as a viable policy instrument for improving mean educational performance or for equalizing outcomes between disadvantaged and advantaged student populations.

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, at approximately between one third to one fifth of the size of our 2SLS estimates. The results of these analyses are presented in Table A.11. The finding that our 2SLS estimates are approximately 5 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 our models with alternate control variable strategies. 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 will 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.12, A.13 and A.14. As expected, results become more precise as covariates are added and efficiency increases, however all results are significant at the same level as in the main results. Point estimates never change by more than .002 standard deviations. These results help support the conclusion that our models represent unconfounded estimates of the impact of expenditures on student performance.

As a second robustness check, we estimate our 2SLS models with higher-order polynomial enrollment terms. Since *Save Harmless* is a function of enrollment, it is necessary to control for the potentially correlated effects of enrollment to render our explanatory variable exogenous. Controlling for higher order polynomial terms should soak up any nonlinear effects of current student enrollment, producing a more robust estimate of *Save Harmless*. These models are provided in Tables A.15, A.16, and A.17. Models for math test scores, both for all students and economically disadvantaged students, are robust to inclusion of a third degree polynomial enrollment term, but fade to significance at the .10 level with inclusion of fourth and fifth degree terms. However English test scores in the economically disadvantaged population, and performance indices in both Math and English are robust to inclusion of polynomial enrollment terms up to degree 7, and possibly higher. In the polynomial models, effect size point estimates are somewhat larger with polynomial enrollment terms than in the main models (for example .08 SD versus .06 SD in math performance). These larger point estimates suggest that if any endogeneity from the relationship between *Save Harmless* and current enrollment exists, it would bias our findings downward – making those presented earlier conservative estimates. Ultimately, the fact that our models are robust to inclusion of between 3 and 7 total polynomial enrollment terms without substantive changes in effect size or significance suggests that district enrollment size is not confounding the results in our main models.

6. Discussion

The findings in our study show clear and compelling evidence that PPE have a positive and significant impact on student outcomes, providing new evidence which contributes to a long-standing debate over the relationship between educational resources and student performance. Leveraging a natural experiment in the context of NYS educational finance mechanics, this study

applies quasi-experimental methods to demonstrate this relationship. While more studies are 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 educational resources do contribute causally to student learning.

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 improve results in high need and low-resource student populations. 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 forcefully refute this argument. 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 produce meaningful impacts on educational performance of approximately .1 standard deviations when applied to struggling and economically disadvantaged students. As achievement gaps 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, (Chaudhary 2009; Jackson, Johnson and

Persico 2016; Papke 2005; Papke 2008) 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 between 61% and 94%, 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. These 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 results 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 large improvements in student achievement. In a state like NY, 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. However, we believe that by performing a rigorous and comprehensive set of falsification tests and robustness checks – none of which raised serious red flags – we have persuasively justified the validity of our inferences. As in all cases of empirical analysis, the results of this study may not provide the definitive statement on the long-standing question of the relationship between educational resources and student performance. Rather, they should be included as another brick in a growing wall of causal evidence that increases educational resources significantly enhances student outcomes.

We offer the most current estimates of the causal impact of PPE, using data as recent as the 2014-15 academic year. We also offer the first estimates 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. This research investigates a novel approach for isolating exogenous variation in educational aid resulting from a school finance reform provision; future research could investigate other state-level finance formulas for natural experiments to shed light on this debate across different 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.

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

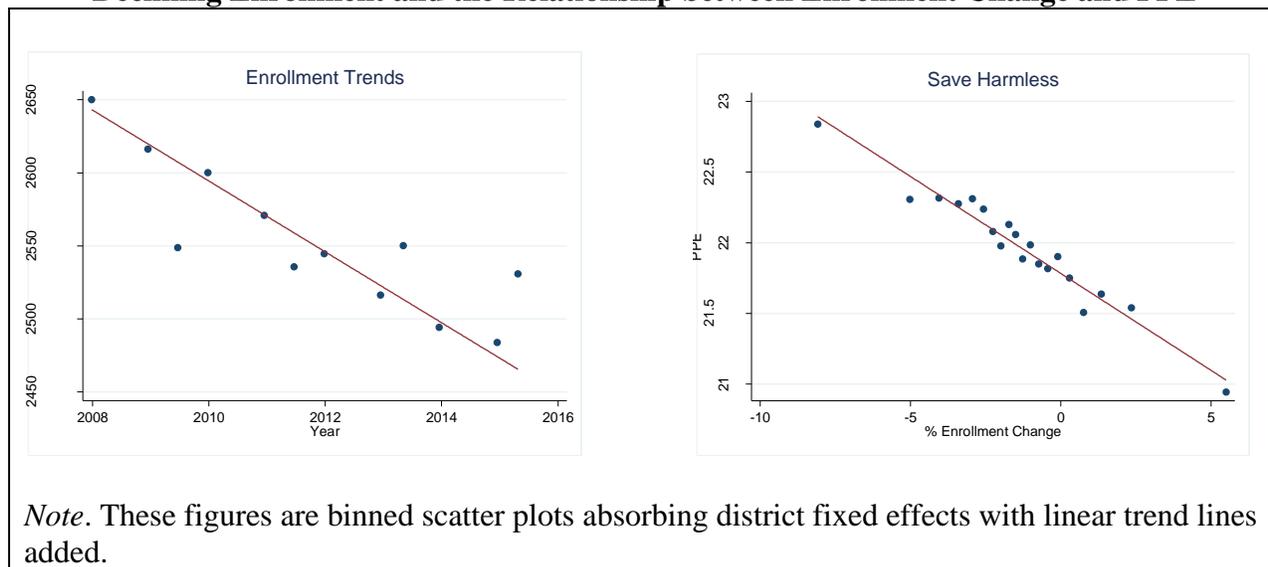
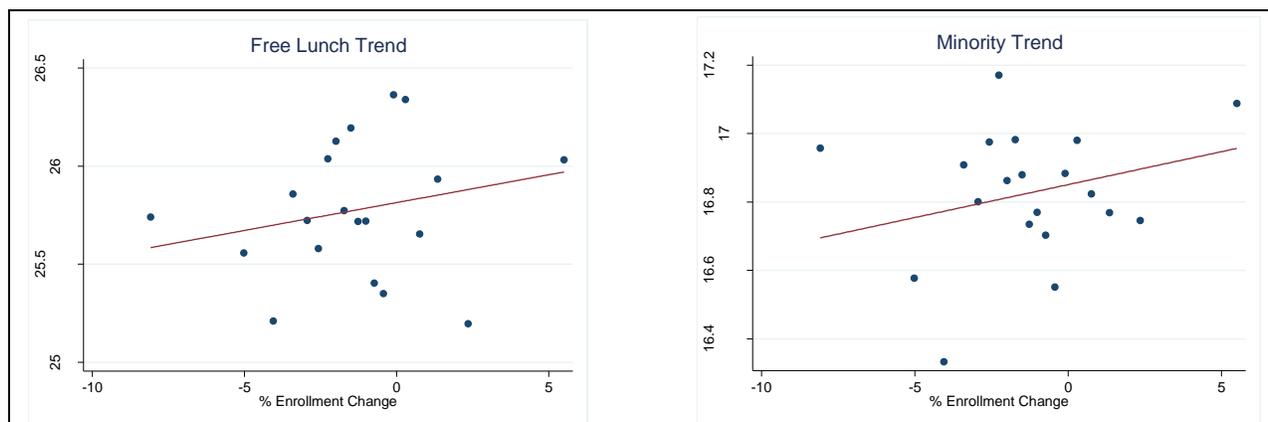
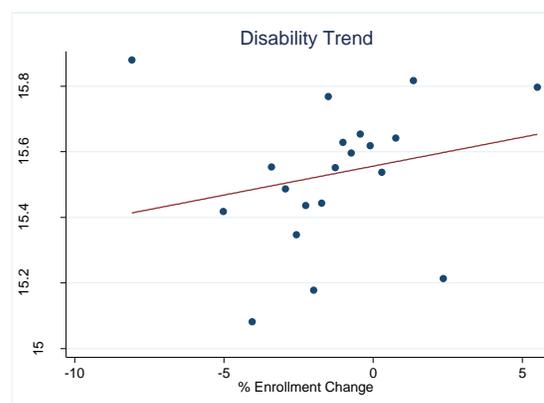
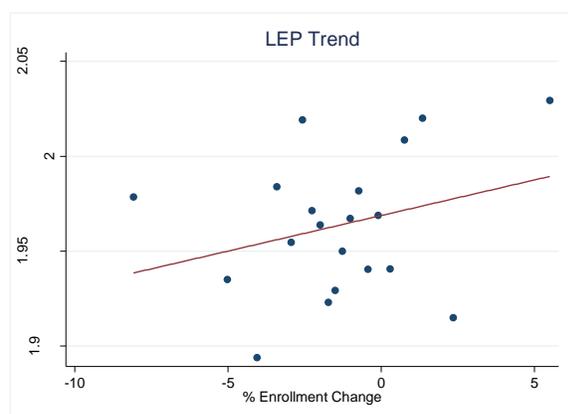


Figure II:
The Relationship between Declining Enrollment and District Demographic Composition





Note. These figures are binned scatter plots absorbing district fixed effects with linear trend lines added.

Variable	Obs	Mean	Std. Dev.	Min	Max
Math Performance Index	5,103	0	1.00	-5.91	2.69
English Performance Index	5,103	0	1.00	-4.31	2.80
Math Score	5,103	0	1.00	-3.58	3.27
English Score	5,103	0	1.00	-3.46	3.37
Math Score (Econ Dis)	4,466	0	1.00	-4.22	4.42
English Score (Econ Dis)	4,466	0	1.00	-3.95	4.43
PPE (in thousands)	5,103	21.97	4.90	12.80	71.70
Student-Teacher Ratio	5,103	11.41	1.68	4.21	25.00
Save Harmless	5,103	1.35	3.21	-19.45	24.84
DCAADM	5,103	2,558.38	3231.07	99.00	43898.00
DCAADM Squared	5,103	1.70E+07	8.81E+07	9.80E+03	1.93E+09
% Free Lunch Eligibility	5,103	25.78	15.87	0.00	95.00
% Minority	5,103	16.83	20.10	0.00	100.00
% LEP	5,103	1.96	4.09	0.00	33.00
% Students with Disabilities	5,103	15.53	4.66	3.21	48.97

	(1)	(2)	(3)	(4)
--	-----	-----	-----	-----

VARIABLES	PPE	PPE	Student	Student
	Model 1	Model 2	Teacher	Teacher
	(\$1000)	(\$1000)	Ratio	Ratio
			Model 1	Model 2
			(Students)	(Students)
Save Harmless	0.1130** (0.0087)	0.1059** (0.0089)	-0.0137** (0.0027)	-0.0128** (0.0027)
DCAADM	-0.0024** (0.0003)	-0.0039** (0.0004)	0.0010** (0.0001)	0.0011** (0.0002)
DCAADM Squared		0.0000** (0.0000)		-0.0000+ (0.0000)
% Free Lunch Eligibility	-0.0069 (0.0054)	-0.0063 (0.0053)	0.0070** (0.0024)	0.0070** (0.0024)
% Minority	-0.0163 (0.0122)	-0.0153 (0.0116)	0.0071+ (0.0043)	0.0069 (0.0042)
% LEP	-0.0775* (0.0389)	-0.0892* (0.0374)	-0.0026 (0.0432)	-0.0011 (0.0432)
% Students with Disabilities	0.0040 (0.0075)	0.0036 (0.0074)	-0.0066* (0.0029)	-0.0065* (0.0029)
Constant	26.8680** (0.7185)	29.6756** (0.8840)	8.6523** (0.3596)	8.3077** (0.4074)
Observations	5,103	5,103	5,103	5,103
R-Squared	0.3755	0.3846	0.1182	0.1194
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	652	652	652	652
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

Table III: F-Test		
VARIABLES	(1)	(2)
	PPE Model 1 (\$1000)	PPE Model 2 (\$1000)
Save Harmless	0.113** (0.0087)	0.106** (0.0089)
F-Statistic	168.81	142.28
Observations	5,103	5,103
R-squared	0.3755	0.3846
Covariates	x	x
Quadratic Enrollment Terms		x
Year FE	x	x
District FE	x	x

Number of District	652	652
Robust standard errors in parentheses		
** p<0.01, * p<0.05, + p<0.1		

	(1)	(2)	(3)	(4)
VARIABLES	Math Model 1 (SD)	Math Model 2 (SD)	English Model 1 (SD)	English Model 2 (SD)
PPE	0.04085** (0.01390)	0.03949** (0.01526)	0.02115 (0.01326)	0.01742 (0.01425)
DCAADM	0.00010 (0.00008)	0.00007 (0.00013)	0.00001 (0.00006)	-0.00009 (0.00010)
DCAADM Squared		0.00000 (0.00000)		0.00000+ (0.00000)
% Free Lunch Eligibility	-0.00508** (0.00134)	-0.00507** (0.00134)	-0.00398** (0.00118)	-0.00398** (0.00118)
% Minority	-0.00719** (0.00269)	-0.00719** (0.00269)	-0.00216 (0.00178)	-0.00216 (0.00178)
% LEP	-0.02698* (0.01204)	-0.02733* (0.01199)	-0.00079 (0.00990)	-0.00177 (0.00993)
% Students with Disabilities	-0.00251 (0.00176)	-0.00251 (0.00176)	-0.00315* (0.00159)	-0.00316* (0.00158)
Constant	-0.85406+ (0.46808)	-0.75713 (0.58770)	-0.33725 (0.42439)	-0.07076 (0.50230)
Observations	5,103	5,103	5,103	5,103
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	652	652	652	652
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

	(1)	(2)	(4)	(5)
VARIABLES	Math Model 1 (SD)	Math Model 2 (SD)	English Model 1 (SD)	English Model 2 (SD)
PPE	0.07430* (0.03033)	0.07469* (0.03419)	0.07193* (0.02970)	0.07301* (0.03327)
DCAADM	0.00011 (0.00014)	0.00012 (0.00024)	0.00016 (0.00012)	0.00018 (0.00021)

DCAADM Squared		-0.00000 (0.00000)		-0.00000 (0.00000)
% Free Lunch Eligibility	-0.00648** (0.00223)	-0.00648** (0.00223)	-0.00450+ (0.00238)	-0.00451+ (0.00238)
% Minority	-0.00617 (0.00463)	-0.00616 (0.00464)	0.01091* (0.00493)	0.01092* (0.00494)
% LEP	-0.04350** (0.01688)	-0.04342** (0.01680)	0.00342 (0.02051)	0.00363 (0.02047)
% Students with Disabilities	-0.00390 (0.00295)	-0.00390 (0.00296)	-0.00325 (0.00288)	-0.00327 (0.00288)
Constant	-1.43306 (0.94733)	-1.45663 (1.21206)	-1.89471* (0.90793)	-1.96006+ (1.15505)
Observations	4,466	4,466	4,466	4,569
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	614	614	614	618
Robust z-statistics in parentheses ** p<0.01, * p<0.05, + p<0.1				

	(1)	(2)	(4)	(5)
VARIABLES	Math Model 1 (SD)	Math Model 2 (SD)	English Model 1 (SD)	English Model 2 (SD)
PPE	0.05507** (0.01817)	0.04852* (0.01933)	0.06608** (0.01459)	0.06302** (0.01563)
DCAADM	0.00023* (0.00011)	0.00006 (0.00015)	0.00032** (0.00008)	0.00024* (0.00012)
DCAADM Squared		0.00000* (0.00000)		0.00000 (0.00000)
% Free Lunch Eligibility	-0.00680** (0.00193)	-0.00678** (0.00192)	-0.00556** (0.00157)	-0.00555** (0.00156)
% Minority	0.00665** (0.00246)	0.00665** (0.00246)	0.00600* (0.00241)	0.00600* (0.00241)
% LEP	0.01727 (0.01724)	0.01555 (0.01654)	0.02014 (0.01352)	0.01934 (0.01340)
% Students with Disabilities	-0.00192 (0.00208)	-0.00193 (0.00208)	-0.00249 (0.00173)	-0.00250 (0.00173)
Constant	-1.73727** (0.64333)	-1.27011+ (0.72734)	-2.20938** (0.49231)	-1.99054** (0.57903)
Observations	5,103	5,103	5,103	5,103
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	652	652	652	652
Robust standard errors in				

parentheses

** p<0.01, * p<0.05, + p<0.1

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 estimate 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.
- 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.
- Chaudhary, Latika, "Education inputs, student performance and school finance reform in Michigan," *Economics of Education Review*, 28 (2009), 90-98.
- 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.
- , "The effects of changes in Michigan's school finance system," *Public Finance Review*, 36 (2008), 456-474.
- 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).
- 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).

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 this 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 performance index and test score dependent variables were missing between 10 and 70 observations each without complete overlap between missing observations. For this reason, we restrict our analysis sample to only those observations for which all variables are not missing. This leads to a total of 97 missing observations for each variable, out of 5,200 possible observations. The test scores which are restricted to economically disadvantaged students contain a more sizable proportion of missing observations. This is because schools without less than 5 economically disadvantaged students per grade do not report these measures. Since this missingness is a function of district poverty and enrollment, which we measure with full information in our model, these observations can be considered missing at random. We perform a similar sample matching procedure for these variables as well.

Variable	Obs	Mean	Std. Dev.	Min	Max
% Students with Disabilities	4892	15.50	4.75	3.21	48.97
% Students with Disabilities (Imputed)	5103	15.53	4.66	3.21	48.97

Table A.2: Conditional Independence (Demographic Characteristics)				
VARIABLES	(1) % Free Lunch Eligibility	(2) % Minority	(3) % LEP	(4) % Student with Disabilities
Save Harmless	-0.00497 (0.0270)	-0.0105 (0.0204)	-0.00292 (0.00250)	-0.0190 (0.0168)
Constant	20.54** (0.162)	14.90** (0.202)	1.826** (0.0242)	15.54** (0.131)
Observations	5,103	5,103	5,103	5,103
R-squared	0.448	0.226	0.038	0.060
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	652	652	652	652
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

Table A.3: Conditional Independence with covariates (Demographic Characteristics)				
VARIABLES	(1) % Free Lunch Eligibility	(2) % Minority	(3) % LEP	(4) % Student with Disabilities
Save Harmless	0.0047 (0.0276)	-0.0216 (0.0235)	0.0023 (0.0030)	-0.0269 (0.0168)
Constant	16.4513** (2.7533)	16.6893** (2.0637)	-0.1208 (0.7469)	17.2805** (1.7986)
Observations	5,103	5,103	5,103	5,103
R-squared	0.4535	0.2396	0.0745	0.0636
Enrollment Term	x	x	x	x
Quadratic Enrollment Term	x	x	x	x
Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Number of District	652	652	652	652
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

Table A.4: Adapted Strict Exogeneity Two-Way FE (Test Scores All Students)				
	(1)	(2)	(3)	(4)
	Math	Math	English	English
VARIABLES	Model 1	Model 2	Model 1	Model 2
	(SD)	(SD)	(SD)	(SD)
Save Harmless (t+1)	0.0004	-0.0004	-0.0011	-0.0012
	(0.0022)	(0.0025)	(0.0021)	(0.0024)
Save Harmless	0.0048**	0.0044*	0.0010	0.0005
	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Enrollment Term	x	x	x	x
Squared Enrollment Term		x		x
Lead Enrollment Terms	x	x	x	x
Covariates	x	x	x	x
Lead Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	4,481	4,481	4,481	4,481
R-Squared	0.0247	0.0258	0.0142	0.0162
Number of District	649	649	649	649
Robust standard errors in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Table A.5: Adapted Strict Exogeneity (Economically Disadvantaged Student Outcomes)				
	(1)	(2)	(3)	(4)
	Math	Math	English	English
VARIABLES	Model 1	Model 2	Model 1	Model 2
	(SD)	(SD)	(SD)	(SD)
Save Harmless (t+1)	0.0012	-0.0020	-0.0033	-0.0049
	(0.0043)	(0.0050)	(0.0040)	(0.0047)
Save Harmless	0.0075*	0.0070*	0.0039	0.0033
	(0.0034)	(0.0035)	(0.0034)	(0.0034)
Enrollment Term	x	x	x	x
Squared Enrollment Term		x		x
Lead Enrollment Terms	x	x	x	x
Covariates	x	x	x	x
Lead Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	3,973	3,973	3,973	3,973
R-Squared	0.0133	0.0144	0.0083	0.0090
Number of District	610	610	610	610
Robust standard errors in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Table A.6: Adapted Strict Exogeneity Two-Way FE (Performance Index)				
	(1)	(2)	(3)	(4)
	Math	Math	English	English
VARIABLES	Model 1	Model 1	Model 1	Model 1
	(SD)	(SD)	(SD)	(SD)
Save Harmless (t+1)	0.0029	0.0008	0.0012	-0.0003
	(0.0032)	(0.0035)	(0.0026)	(0.0029)
Save Harmless	0.0064*	0.0054*	0.0069**	0.0061**
	(0.0025)	(0.0025)	(0.0019)	(0.0019)
Enrollment Term	x	x	x	x
Squared Enrollment Term		x		x
Lead Enrollment Terms	x	x	x	x
Covariates	x	x	x	x
Lead Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	4,481	4,481	4,481	4,481
R-Squared	0.0283	0.0334	0.0281	0.0329
Number of District	649	649	649	649
Robust standard errors in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Table A.7: Strict Exogeneity 2SLS (Test Scores All Students)				
	(1)	(2)	(3)	(4)
	Math	Math	English	English
VARIABLES	Model 1	Model 2	Model 1	Model 2
	(SD)	(SD)	(SD)	(SD)
PPE (t+1)	-0.0057	-0.0107	-0.0093	-0.0093
	(0.0135)	(0.0154)	(0.0132)	(0.0152)
Enrollment Term	x	x	x	x
Squared Enrollment Term		x		x
Lead Enrollment Terms	x	x	x	x
Covariates	x	x	x	x
Lead Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	4,481	4,481	4,481	4,481
Number of District	649	649	649	649
Robust standard errors in parentheses				
** p<0.01, * p<0.05, + p<0.1				

Table A.8: Strict Exogeneity 2SLS (Test Scores Economically Disadvantaged Students)				
VARIABLES	(1)	(2)	(3)	(4)
	Math Model 1 (SD)	Math Model 2 (SD)	English Model 1 (SD)	English Model 2 (SD)
PPE (t+1)	-0.0077 (0.0319)	-0.0323 (0.0374)	-0.0354 (0.0291)	-0.0465 (0.0344)
Enrollment Term	x	x	x	x
Squared Enrollment Term		x		x
Lead Enrollment Terms	x	x	x	x
Covariates	x	x	x	x
Lead Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	3,973	3,973	3,973	3,973
Number of District	610	610	610	610
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

Table A.9: Strict Exogeneity 2SLS (Performance Index)				
VARIABLES	(1)	(2)	(3)	(4)
	Math Model 1 (SD)	Math Model 2 (SD)	English Model 1 (SD)	English Model 2 (SD)
PPE (t+1)	0.0087 (0.0196)	-0.0045 (0.0212)	-0.0040 (0.0158)	-0.0125 (0.0178)
Enrollment Term	x	x	x	x
Squared Enrollment Term		x		x
Lead Enrollment Terms	x	x	x	x
Covariates	x	x	x	x
Lead Covariates	x	x	x	x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	4,481	4,481	4,481	4,481
Number of District	649	649	649	649
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

VARIABLES	(1) PPE (\$1000) (Urban High Need)	(2) PPE (\$1000) (Rural High Need)	(3) PPE (\$1000) (Medium Need)	(4) PPE (\$1000) (Low Need)	(5) PPE (\$1000) (Untreated)
Save Harmless	0.0644** (0.0164)	0.0547** (0.0111)	0.0969** (0.0148)	0.137** (0.0258)	0.121** (0.0219)
Enrollment Term	x	x	x	x	x
Squared Enrollment Term	x	x	x	x	x
Covariates	x	x	x	x	x
Year FE	x	x	x	x	x
District FE	x	x	x	x	x
Observations	359	1,217	2,628	898	1,495
R-squared	0.409	0.362	0.479	0.507	0.213
Number of District	53	174	363	126	580
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1					

VARIABLES	(1) Math Scores (SD)	(2) English Scores (SD)	(3) Math Scores (SD)(Econ Dis)	(4) English Scores (SD) (Econ Dis)	(5) Math PI (SD)	(6) English PI (SD)
PPE	0.01537** (0.00402)	0.01106** (0.00345)	0.02246** (0.00589)	0.0179** (0.00616)	0.01090* (0.00453)	0.0119** (0.00403)
DCAADM	0.00003 (0.00007)	-0.00002 (0.00005)	-0.00003 (0.00010)	5.08e-06 (7.93e-05)	0.00010 (0.00008)	0.000157** (5.91e-05)
% Free Lunch Eligibility	-0.00525** (0.00131)	-0.00405** (0.00116)	-0.00675** (0.00215)	-0.00478* (0.00228)	-0.00710** (0.00186)	-0.00592** (0.00145)
% Minority	-0.00768** (0.00284)	-0.00236 (0.00179)	-0.00764+ (0.00460)	0.00937* (0.00468)	0.00581* (0.00238)	0.00496* (0.00210)
% LEP	-0.02872* (0.01210)	-0.00148 (0.00982)	-0.04571** (0.01696)	0.00111 (0.0201)	0.01424 (0.01710)	0.0164 (0.0131)
% Students with Disabilities	-0.00247	-0.00313*	-0.00370	-0.00305	-0.00185	-0.00240

Constant	(0.00176) -0.12222 (0.21845)	(0.00159) -0.04728 (0.16738)	(0.00297) 0.02072 (0.33599)	(0.00291) -0.381 (0.298)	(0.00208) -0.46858+ (0.27998)	(0.00167) -0.654** (0.209)
Observations	5,103	5,103	4,466	4,466	5,103	5,103
R-Squared	0.02864	0.01619	0.01792	0.008	0.01550	0.021
Year FE	x	x	x	x	x	x
District FE	x	x	x	x	x	x
Number of District	652	652	614	614	652	652
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1						

VARIABLES	(1)	(2)	(3)	(4)
	Math (SD)	Math (SD)	English (SD)	English (SD)
PPE	0.04093** (0.01555)	0.03949** (0.01526)	0.01845 (0.01440)	0.01742 (0.01425)
Enrollment Terms	x	x	x	x
Squared Enrollment Terms	x	x	x	x
Covariates		x		x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	5,103	5,103	5,103	5,103
Number of District	652	652	652	652
Robust z-statistics in parentheses ** p<0.01, * p<0.05, + p<0.1				

VARIABLES	(1)	(2)	(3)	(4)
	Math (SD)	Math (SD)	English (SD)	English (SD)
PPE	0.07278* (0.03488)	0.07469* (0.03419)	0.07373* (0.03359)	0.07301* (0.03327)
Enrollment Term	x	x	x	x
Squared Enrollment Term	x	x	x	x
Covariates		x		x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	4,466	4,466	4,466	4,466
Number of District	614	614	614	614
Robust standard errors in parentheses ** p<0.01, * p<0.05, + p<0.1				

VARIABLES	(1) Math (SD)	(2) Math (SD)	(3) English (SD)	(4) English (SD)
PPE	0.04759* (0.01965)	0.04852* (0.01933)	0.06248** (0.01571)	0.06302** (0.01563)
Enrollment Term	x	x	x	x
Squared Enrollment Term	x	x	x	x
Covariates		x		x
Year FE	x	x	x	x
District FE	x	x	x	x
Observations	5,103	5,103	5,103	5,103
Number of District	652	652	652	652

Robust standard errors in parentheses
 ** p<0.01, * p<0.05, + p<0.1

VARIABLES	(1) Math (SD)	(2) Math (SD)	(3) Math (SD)	(4) English (SD)	(5) English (SD)	(6) English (SD)
PPE	0.03824* (0.01774)	0.03351+ (0.01985)	0.03215 (0.02114)	0.01369 (0.01611)	0.01293 (0.01797)	0.01266 (0.01911)
Enrollment Term	x	x	x	x	x	x
Enrollment Term ^2	x	x	x	x	x	x
Enrollment Term ^3	x	x	x	x	x	x
Enrollment Term ^4		x	x		x	x
Enrollment Term ^5			x			x
Covariates	x	x	x	x	x	x
Year FE	x	x	x	x	x	x
District FE	x	x	x	x	x	x
Observations	5,103	5,103	5,103	5,103	5,103	5,103
Number of District	652	652	652	652	652	652

Robust standard errors in parentheses
 ** p<0.01, * p<0.05, + p<0.1

VARIABLES	(1) Math (SD)	(2) Math (SD)	(3) Math (SD)	(4) English (SD)	(5) English (SD)	(6) English (SD)
PPE	0.08245* (0.04161)	0.08068+ (0.04857)	0.08456 (0.05256)	0.11526* (0.05066)	0.12707* (0.05528)	0.14123* (0.06123)
Enrollment Term	x	x	x	x	x	x

Enrollment Term ^2	x	x	x	x	x	x
Enrollment Term ^3	x	x	x	x	x	x
Enrollment Term ^4		x	x	x	x	x
Enrollment Term ^5			x	x	x	x
Enrollment Term ^6					x	x
Enrollment Term ^7						x
Covariates	x	x	x	x	x	x
Year FE	x	x	x	x	x	x
District FE	x	x	x	x	x	x
Observations	4,466	4,466	4,466	4,466	4,466	4,466
Number of District	614	614	614	614	614	614
Robust standard errors in parentheses						
** p<0.01, * p<0.05, + p<0.1						

Table A.17: Polynomial Terms 2SLS (Performance Index)						
VARIABLES	(1) Math (SD)	(2) Math (SD)	(3) Math (SD)	(4) English (SD)	(5) English (SD)	(6) English (SD)
PPE	0.07782** (0.02709)	0.08408** (0.02903)	0.09283** (0.03113)	0.08493** (0.02219)	0.08849** (0.02391)	0.09365** (0.02577)
Enrollment Term	x	x	x	x	x	x
Enrollment Term ^2	x	x	x	x	x	x
Enrollment Term ^3	x	x	x	x	x	x
Enrollment Term ^4	x	x	x	x	x	x
Enrollment Term ^5	x	x	x	x	x	x
Enrollment Term ^6		x	x		x	x
Enrollment Term ^7			x			x
Covariates	x	x	x	x	x	x
Year FE	x	x	x	x	x	x
District FE	x	x	x	x	x	x
Observations	5,103	5,103	5,103	5,103	5,103	5,103
Number of District	652	652	652	652	652	652
Robust standard errors in parentheses						
** p<0.01, * p<0.05, + p<0.1						

