EDUCATION RECOVERY
SCORECARD

Federal Pandemic Relief and Academic Recovery

Dan Dewey, *Harvard University*
Erin Fahle, *Stanford University*
Thomas J. Kane, *Harvard University*
Sean F. Reardon, *Stanford University*
Douglas O. Staiger, *Dartmouth College*

Center for Education Policy Research
HARVARD UNIVERSITY

The Educational Opportunity Project
at Stanford University

I. Introduction

Since the beginning of the War on Poverty, social scientists have been investigating the link between unequal school spending and gaps in achievement by race and income (see, e.g., Coleman et al., 1966; Handel & Hanushek 2023; Jackson & Mackevicius 2024). Over the decades, research has shown that certain interventions—such as pre-school, reduced class size in early grades, summer learning and high-dosage tutoring—do have positive causal impacts on student achievement (Harris 2009; Nickow, Oreopolous, Quan 2020). So, while there is no longer a question of whether more money for schools could improve student outcomes, there is still a question of magnitude: when given additional funds, how effectively do school boards and district leaders spend them to improve student outcomes and by how much?

The recent influx of federal pandemic relief provides a fresh opportunity to test the relationship between spending and achievement. During 2020 and 2021, the federal government approved three aid packages totaling nearly $190 billion for elementary and secondary schools, ninety percent of which was provided directly to local school districts with little federal or state oversight. The aid was not based on student learning losses nor on the amount of time students were out of school. In fact, the final package of aid was approved in March 2021, while many schools were still operating remotely and before the magnitude of losses were even known. Instead, the grants were provided proportional to each districts' funding under the federal Title I program in fiscal years 2019 and 2020.

Because Title I funding is based on local poverty rates for school-age children, the federal relief went primarily to higher poverty districts. Therefore, our primary empirical challenge is distinguishing between the effect of the additional spending and the effect of community poverty. We take two different approaches to accomplish this.

First, we measure the relationship between federal pandemic relief spending per student and the change in average student achievement between spring 2022 and 2023, while statistically controlling for a variety of district characteristics, such as student demographics and two different measures of community poverty (the share of school-age children living in the district meeting Title I eligibility criteria and the share of public school students eligible for federal subsidized lunches).

Our estimate of the impact of the federal relief spending is in line with a recent meta-analysis of pre-pandemic studies on the effect of education spending (Jackson and Mackevicius, 2024). While Jackson and Mackevicius (2024) report an impact of .0079 SD per $1000 increase in spending per student, we find a .0086 SD rise in math and a .0049 SD rise in reading scores per $1000 of federal relief spending. After controlling for federal relief dollars spent and other district characteristics, we find no relationship between federal relief dollars that had not yet been spent by May-June 2023 and achievement growth during 2022-23. We take the latter as a form of placebo test, since dollars not yet spent should have no relationship to achievement growth.

We investigate the role of three specific sources of variation in district spending. To test whether our findings are driven by the timing of district expenditures, we use total federal relief per student as an instrumental variable for district spending (essentially assuming districts spent the same share of funds during the 2022-23 school year) and find that the districts which received larger allocations increased more and that the implied effect of dollars spent was similar to our earlier estimates.

We also focus on the differences in relief spending driven by differences in state Title I funding formulae.

The vast majority of the pandemic relief ($175 billion of $190 billion) was based on Title I allocations during fiscal year 2020. Since the pandemic relief amounted to more than 10 times the federal Title I funding in that year, the funding magnified pre-existing anomalies in the Title I formula. For instance, if a district in one state received $500 more Title I dollars per student than a similar district in another state during fiscal year 2020, they would have received roughly $5,000 more per student in federal pandemic aid. When we focus on the differences in spending attributable to differences in state Title I formulae, we also see positive impacts consistent with prior research.

Finally, we focus on differences in relief spending driven by seemingly random fluctuations in district poverty estimates provided by the U.S. Bureau of the Census. Although the latter are the most plausibly exogenous source of variation we use, these estimates are also the most imprecise: while not statistically distinguishable from zero, they are also not distinguishable from Jackson and Mackevicius (2024) or our other estimates.

As a second test, we identify high poverty districts with similar trends in achievement between 2016 and 2022, but which received differing amounts of federal aid per student. While our first strategy prioritized controlling for measurable district characteristics, the second strategy prioritizes controlling for any unmeasured factors underlying district achievement trends. Of 785 districts with more than 70 percent of students receiving federal lunch subsidies in our data, we identify 149 districts which received unusually large ESSER allocations per student (more than $8200 per student) and compare them with combinations of districts with unusually small ESSER allocations (less than $4600 per student), but with similar trends in achievement between 2016 and 2022. We find that the average scores for students in high-grant districts increased by .055 SD more in math and .047 SD more in reading between 2022 and 2023. Given a difference in spending between the two groups of $2820 per student, our second set of estimates imply an impact per dollar spent of roughly .02 standard deviations per $1000 spent, roughly double our first set of estimates, though the confidence intervals on these estimates are large and do not rule out effects of the same magnitude as Jackson and Mackevicius (2024).

In sum, our results imply that the federal pandemic relief contributed to academic recovery during the 2022-23 school year, and that the impacts were in line with what would have been expected from prior research. Because the federal relief dollars were disproportionately targeted at low-income districts, they are contributing to narrowing the gaps which widened during the pandemic. We close by discussing ways that any additional aid—such as from states—could be structured to yield larger impacts on student achievement and close the remaining gaps.

II. Literature Review

In a recent review of the literature, Jackson and Mackevicius (2024) identify 32 studies employing quasi-experimental designs to study the effect of school spending on student outcomes. The authors excluded cross-sectional correlational analyses and limited themselves to studies using more plausibly exogenous design-based controls—regression discontinuity, difference-in-differences, event studies or instrumental variables. For the subset of studies using test scores as an outcome, their meta-analysis implied an average impact of .0316 standard deviations per $1000 increase in spending over 4 years. Since we are looking at the impact of higher spending within a single academic year, 2022-23, we divide the estimated impact by four, yielding an estimate of .0079 standard deviations per $1000 in a single year.

Even if .0079 SD per $1000 is the mean impact of a $1000 increase in general revenues, a number of targeted academic interventions have been shown to have greater impact per dollar spent. For instance,

in summarizing the research from the Tennessee class size experiment, Harris (2009) concluded that the Tennessee classroom size experiment, which reduced K-3 class size from an average of 22 students per teacher to 15 students per teacher, generated average gains of .063 SD per $1000 spent.[1] Harris (2009) also summarized the evidence from an evaluation of the Success for All whole school reform model (Borman et al. 2007), and concluded that the implied impact was .085 SD per $1000 spent (when converted from 2007 to 2022 dollars). More recently, Guryan et al. (2023) report on the effect of two high-dosage tutoring programs for secondary students, finding a pooled impact of .28 standard deviations for those participating in a program costing $3500 per student. That would have translated into .08 standard deviations per $1000 spent—roughly 10 times the impact of an increase in general funds reported by Jackson and Mackevicius (2024).

Thus, even if more spending is related to higher student outcomes, there may be opportunities to increase the bang for the buck. It matters not just whether a specific intervention leads to improvements, but how large the effects are per dollars spent.

The federal pandemic relief dollars were not intended solely for academic recovery. Indeed, the American Rescue Plan only required districts to spend a minimum of 20 percent on academic recovery. For instance, many districts purchased masks for students and teachers and distributed food and devices. Yet, in a typical year, districts spend a much larger share of their revenues on instruction. Of the $769 billion in annual expenditures for public K-12 education in 2018-19 (the year before the pandemic), the National Center for Education Statistics reports that public school districts spent a larger share, 52 percent, on instruction.[2] For this reason, one might have expected the impact per dollar spent to have been lower than found in the prior research.

III. Data

To measure the impact of the federal pandemic relief dollars, we use data on test scores, district-level poverty rates, Title I funding, federal pandemic relief funding, district characteristics and the percent of the 2020-21 school year that the district was operating remotely or a hybrid of remote and in-person instruction. We obtain these data from multiple sources, describing each source of data below.

Test score data: Our outcome measures are estimates of district average test scores, based on state standardized tests in math and reading in grades 3 through 8, from 2016-2019 and in 2022 and 2023.[3] Most states do not report average test scores, however. Instead, they report the proportion of students in a district who score at each of several state-defined proficiency levels. Assuming that the raw scores in each district are normally distributed, we use heteroskedastic ordered probit models to estimate the mean score in each district from the counts in each proficiency category (for details on this method, see Reardon, Kalogrides, & Ho, 2021; Reardon, Shear, Castellano, & Ho, 2017; Shear & Reardon, 2021). The method yields estimates of district average scores that are comparable among districts in the same state-year-grade. However, state math and reading tests, and the scale in which scores are reported, differ among states and grades; in some cases, they vary over time within the same state-grade. To put the scores from each state's test on the same scale, we link the state test scores to the National Assessment of Educational Progress (NAEP) test scale,

---

[1]   Table 3 in Harris (2009) reports a short term cost effectiveness ratio for participating students of .086 standard deviations per $1000 of 2007 dollars. We converted to 2022 dollars using the CPI-U, dividing by 1.37.
[2]   https://nces.ed.gov/programs/digest/d23/tables/dt23_236.10.asp?current=yes
[3]   We use test score data from the Spring of 2016-2019 and from Spring 2022 and Spring 2023. Data for the years 2016-2019 is available for all states from EDFacts. We collected 2022 and 2023 test score results from state department of education websites.

using methods described in Reardon, Kalogrides, and Ho (2021). We then standardize the scores within each grade relative to the national student-level distribution of scores in 2019. Following this, we estimate average district scores that are comparable over time and across states within each grade. To pool the average scores across grades within a district, we regress average scores on a linear grade term (centered at grade 5.5, so that our estimates apply to the middle grade of the data even in cases where data is missing for one or mode grades in a district). Details on the construction of the estimates are provided in Fahle et al (2024b).

Because many states waived testing requirements in 2020 and 2021 due to the pandemic, we measure pandemic losses by comparing achievement in Spring 2019 and Spring 2022. Thus, we measure academic recovery as the change in average test scores from Spring 2022 to Spring 2023, even though many districts will have started recovery efforts during the 2021-22 school year.

Because the NAEP test was not administered in 2023 (and the 2024 results are not yet available), we cannot yet put the 2023 proficiency categories on the NAEP scale. Instead, we contacted the department of education in each state and identified the subset of states who reported that their tests and proficiency thresholds remained unchanged between 2022 and 2023 (29 of the 42 states included in our analysis from 2019 to 2022). We then use our estimates of the 2022 proficiency thresholds to estimate 2023 district scores in that subset of states.[4]

Because students in any given state take the same tests in a given grade and year, a comparison of test score changes between two districts in the same state does not depend on our method of linking test scores to a common scale across states and over time (using NAEP though 2022 and using the stability of state proficiency thresholds from 2022-2023). But comparisons between districts in different states do rely on the accuracy of the linking. The linking is not exact, both because the state NAEP estimates used for linking contain sampling error, and because, even when a state does not change its test from one year to the next, the tests are not identical—they contain different items—and so their scales are not perfectly identical. As a result, within-state comparisons of changes (including estimates from state fixed effects models) are less error-prone than between-state comparisons. That said, Reardon, Kalogrides, and Ho (2021) demonstrated that the NAEP-based linking yield valid comparisons across states, albeit with some uncertainty.

Because we control for local area poverty rates from the Census Bureau, we focus on traditional school districts which have geographic boundaries. Although some charter schools are administered by their local district and, thus, are included with the district estimates, we do not include independent charter schools which constitute their own local education agency.

Our final analysis sample consists of 5812 districts in 29 states. Table 1 reports the characteristics of our final analysis sample against the full set of traditional public school districts in the 42 states whose achievement we measured in 2019 and 2022.

---

[4]   When the 2024 NAEP results become available in early 2025, we will update our 2023 estimates, by interpolating between the 2022 and 2024 NAEP-based estimates.

TABLE 1. SAMPLE MEANS AND STANDARD DEVIATIONS FOR ANALYSIS SAMPLE VS. FULL SAMPLE

|	Final Analysis Sample (29 States)		Full Sample (42 States)	
	Mean	SD	Mean	SD
Achievement Outcomes:				
2019-22 Change in Math	-0.149	0.114	-0.154	0.119
2019-22 Change in Reading	-0.095	0.091	-0.092	0.088
2022-23 Change in Math	0.046	0.052	0.046	0.052
2022-23 Change in Reading	0.019	0.061	0.019	0.061
Demographics:				
% Black (2022)	14.9%	18.9%	15.2%	17.7%
% Hispanic (2022)	25.0%	23.8%	28.2%	24.8%
% Free Lunch (2022)	47.9%	21.8%	48.3%	21.0%
Log Total Enrollment (2022)	9.270	1.484	9.475	1.602
Percent of Population Aged 5-17 Who Met Title I Formula (2020)	16.6%	9.4%	16.9%	9.2%
Percent Remote/Hybrid:				
Percent of 2020-21 SY in Remote Instruction	30.2%	29.4%	24.4%	28.4%
Percent of 2020-21 SY in Hybrid Instruction	43.4%	27.4%	37.1%	28.3%
ESSER II/ARP:				
ESSER II/ARP Dollars Allocated Per Student	$3,167	$2,655	$3,129	$2,578
ESSER II/ARP Dollars Spent Per Student	$1,961	$1,660	$1,961	$1,613
Number of Districts	5,812		9,424	
Number of Students	26,107,534		39,345,566	

The analysis sample is similar to the full sample. The achievement losses in math and reading (.149 and .095 SD respectively) were comparable to those in the full sample of 42 states (.154 and .092 SD respectively). The students were similar demographically and the amount of aid districts received under ESSER II and the American Rescue Plan were similar ($3167 vs. $3129 per student). The primary difference is that the average district in the analysis sample spent a somewhat larger share of the 2021 school year in remote and hybrid instruction (30 and 43 percent respectively vs. 24 and 37 percent in the full set of states.)

Estimates of the Number and Percentage of Title I Eligible Children by District: To identify sources of variation in Title I grants, we use estimates of the number and percentage of eligible children living in more than 13,000 geographical school districts for fiscal years 2013 through 2023. Although the number of eligible children includes children in other categories (such as the number of foster children, neglected and delinquent children and students attending Bureau of Indian Education schools in the district), the primary driver of Title I allocations is the estimated number and percentage of school age children (those aged 5-17) living in poverty.

The Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program provides the poverty estimates annually, using administrative data (from the Supplemental Nutrition Assistance Program (SNAP) and geographically-linked federal tax records) as well as survey data from the American Community Survey.

Because many school districts are quite small, the poverty rate estimates are subject to substantial measurement error. As an illustration, Figure 1 plots the annual changes in the percent of 5–17-year-olds estimated to meet Title I eligibility between fiscal year 2013 and 2023 (on the vertical axis). On the horizontal axis, we plot the number of 5–17-year-olds estimated to live within the district boundaries. There are obviously large swings in the share of 5-17 year-olds considered to be eligible for Title I (sometimes larger than 25 percentage points in the smaller districts).

FIGURE 1: SINGLE-YEAR CHANGES IN PERCENT OF 5-17 YEAR OLDS TITLE I ELIGIBLE BY DISTRICT SIZE, FY 2013-2023
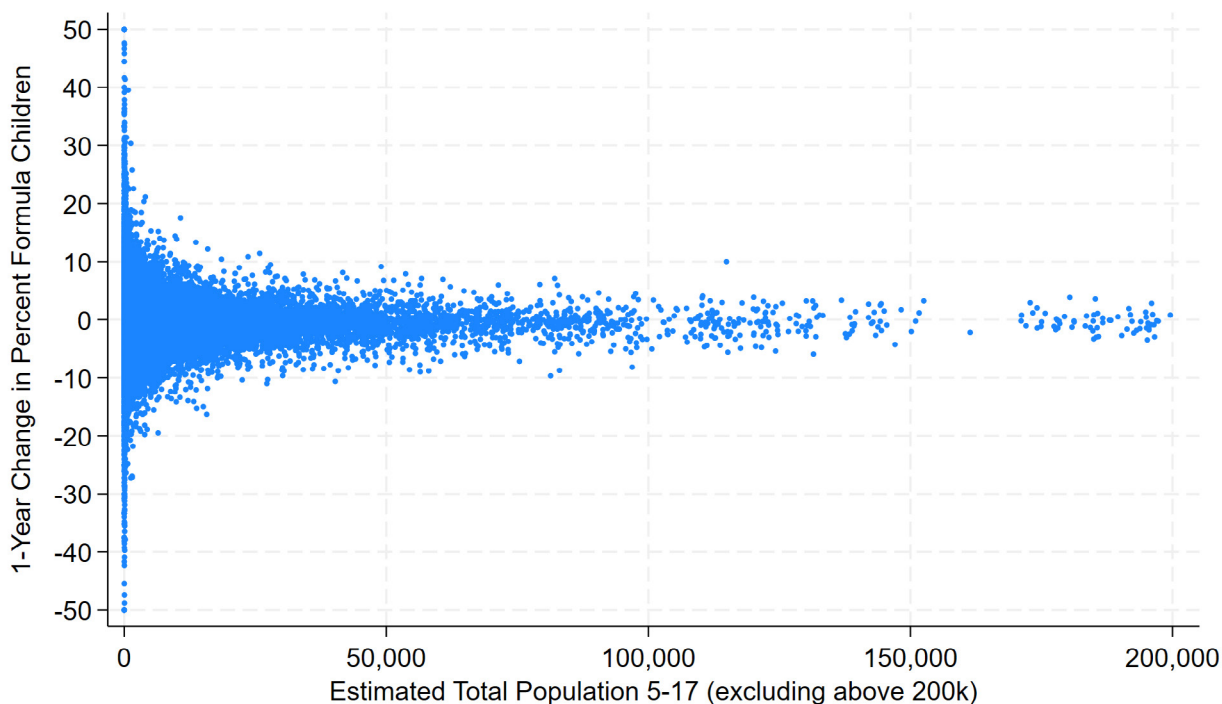
Note: We exclude from this figure districts with more than 200,000 5–17-year-olds and districts with greater than a 50 percentage point change in the proportion of children meeting the Title I eligibility definition.

Figure 2 portrays the trend over time in estimated eligibility rates in one district, Gary, Indiana, between 2013 and 2023. The dotted line portrays the trend in the estimates. The rate increased sharply in Fiscal Year 2020, which was fortunate for residents of Gary since 2020 was the year upon which the bulk of the federal pandemic relief dollars were based. As a result, Gary's Title I allocation increased by 23 percent between 2019 and 2020, from $11.4 million to $14.0 million. Given that the ESSER II and American Rescue Plan allocations were based on the 2020 estimates and provided roughly 10 times the amount of dollars, the change resulted in a $26 million windfall for the Gary community.

However, not all of the fluctuation depicted in Figure 2 appears to be random. Eligibility rates have been drifting down in Gary since 2013. That means that it would be problematic to use any change over time in eligibility rates as an instrument for federal pandemic relief. For instance, if eligibility rates had been higher

in 2020 than in 2023, it would mean that a district received a larger ESSER grant than other districts with the same 2023 eligibility rate. But it might also mean that the community was becoming richer (less poor), with an expected upward trend in achievement. Due to the underlying trend in true eligibility, the estimated effect of spending could be biased upward.

Thus, we attempted to isolate the seemingly random fluctuation in eligibility rates due to measurement error from the systemic, structural trend. To do so, we used a Hodrick-Prescott time series filter (Hodrick and Prescott, 1997) to separate out the structural trend in the time series (portrayed by the red curve) from the actual (the dotted line). Changes in eligibility along the structural trend (the red line) could be directly related to achievement and thus could lead to bias. But if the fluctuation in eligibility rates around the trend line truly were due to measurement error in the American Community Survey and other sources, they would be related to ESSER grants but should not be directly related to achievement. To find the variation ESSER funding which was truly "as good as random", we use the difference between what the district actually received in FY 2020 vs. the amount the district would have received if the percentage of eligible children had followed the structural trend (the red line).

FIGURE 2. TREND IN PERCENT FORMULA CHILDREN IN GARY, INDIANA, 2013-2023
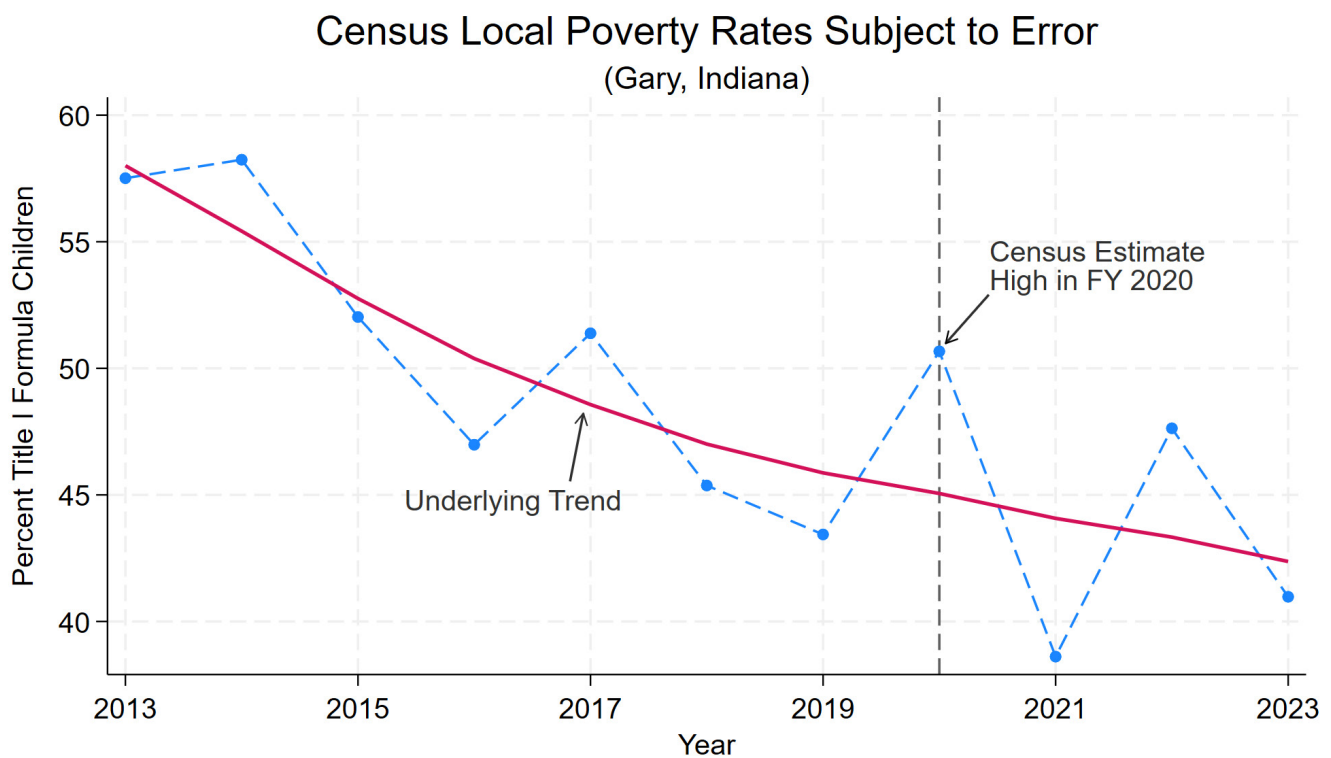
In Figure 3, we plot the difference in grants per child aged 5-17 due to the seemingly random fluctuations in eligibility rates around the structural trend (along the vertical axis). We plot the differences by the number of school-age children (ages 5-17) in each district (on the horizontal axis). Note that much of the variation is isolated to small districts, with fewer than 5,000 school age children. Moreover, because hold harmless provisions mute the effect of downward fluctuations, most of the points are positive. Although the results are imprecise, we use this variation as an instrumental variable in our empirical analysis below.

FIGURE 3. VARIATION IN TITLE I ALLOCATIONS DUE TO FLUCTUATIONS IN DISTRICT ELIGIBILTY RATES
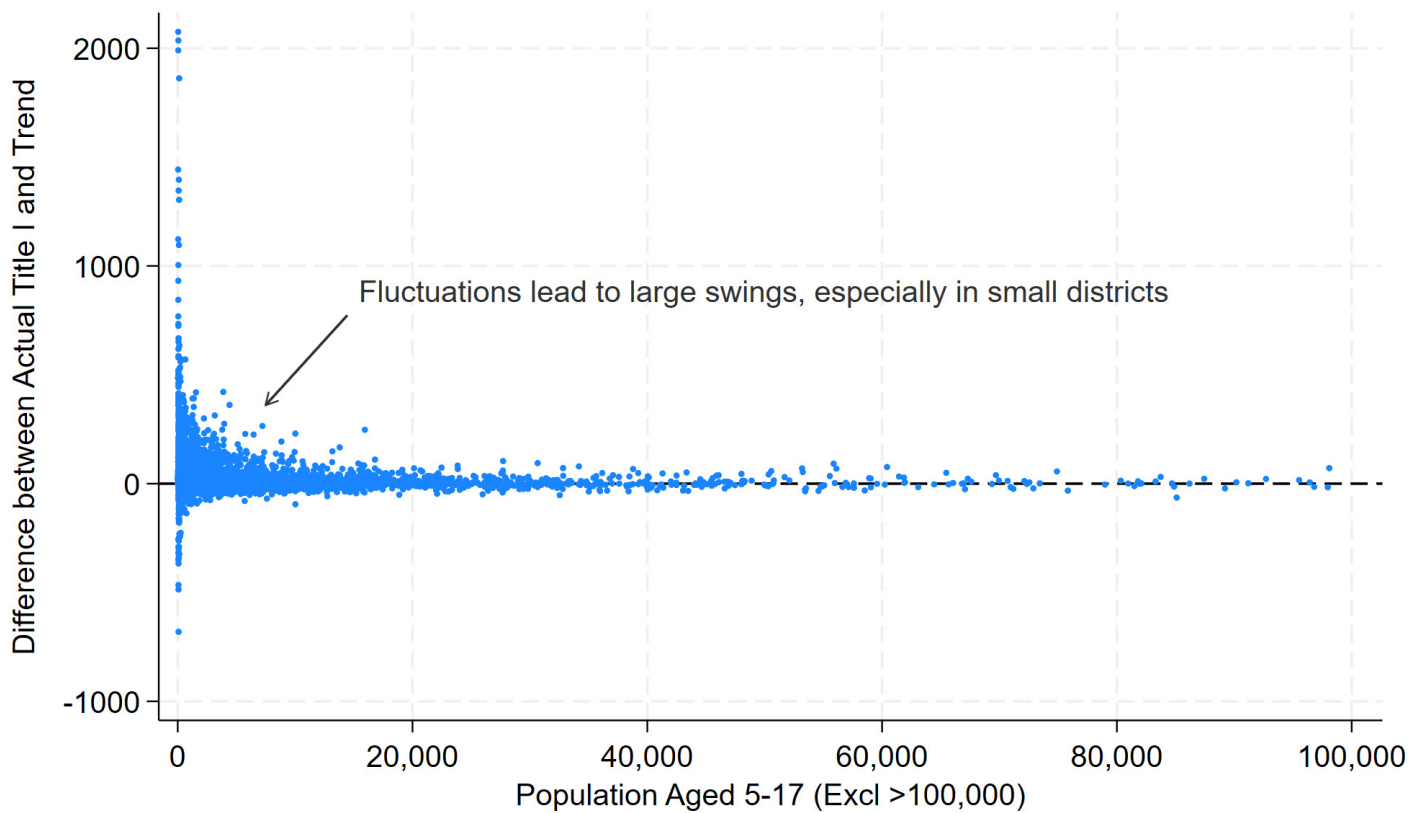
Note: For each district, we estimated a smoothed trend in the percent of students meeting the Title I eligibility criteria between 2013 and 2023, using the Hodrick-Prescott time series filter. We then simulated what each district would have received in Title I in each year had their percent eligible children followed the trend. The above is the difference between the Title I allocation based on the trend and actual Title I per student in FY 2020.

Title I Program: In addition to the data on the number and proportion of children estimated to be eligible in each district, we received data on the Title I allocations for fiscal years 2013 through 2023. States adjust the Title I allocations calculated based on the population living within school district boundaries, redistributing those dollars to the traditional public schools and to charter districts based on estimates of the number of poor children they educate. In addition, a state can request approval to use alternate data (e.g., school lunch data) to redistribute the federal department's allocations for small districts (defined in ESEA as having a total Census population below 20,000). Currently, 11 states have the federal department's approval to redistribute their small LEAs' allocations. We obtained data on state-adjusted Title I allocations for districts from ED Data Express and used them to allocate Title I aid proportionately.

Federal Pandemic Relief Spending: The first package of federal pandemic relief for schools ($13.2 billion) had to be obligated by districts by the end of September 2022. Because we are investigating achievement gains between Spring 2022 and Spring 2023, we focus on the second two packages, ESSER II ($54 billion) and the American Rescue Plan or ESSER III ($122 billion). We received data from a private company, Burbio, on the amount of ESSER II/III funding each district received as well as the amount of money districts reported spending as of specific dates (although the dates varied by state). For 5 states, we supplemented

the Burbio data with data from the website Edunomics.org, including ESSER II allocations. Since ESSER II and ESSER III were distributed proportionally within each state using the same rules, we impute missing ESSER II allocations using the district's ESSER III allocation and the ratio of the state's total ESSER II and III allocations. We used the total amount districts reported having spent from ESSER II/III in the earliest available reporting window in May 2023 or later. Since ESSER II had to obligated by September 2023 we assume all ESSER II allocation was spent between Spring 2022 and Spring 2023. Due to ESSER III data availability, 50% of districts have reporting windows before August 2023, with 80% before December 2023. This means our spending measures likely overstate the amount spent between Spring 2022 and Spring 2023, as we may be capturing spending either prior to Spring 2022 or after Spring 2023, thereby making our estimates more conservative. We find that controlling flexibly for the date spent does not impact our regression results.

District Federal Subsidized Lunch Eligibility: We use district free and reduced-price lunch eligibility (FRLE) rate estimates computed from the Longitudinal Imputed School Dataset (LISD) (Reardon et al, 2024). The LISD includes reported or imputed school-level FRLE rates for each public school in the U.S. for the school years 1998-99 through 2022-23. The FRLE rates are based on data from the Common Core of Data (CCD). In the CCD, however, FRLE rates are missing in some cases and clearly erroneous in others; in addition, schools that are identified as using providing meals through the Community Eligibility Provision (CEP) often have FRLE rates reported as 100%, even though not all students in the school are individually eligible for FRL. In cases of missing or erroneous FRLE values, the reported values are replaced by imputed FRLE rates. The imputation uses data from the same school in other years, and uses FRLE rates, racial composition, the proportion of students identified as economically disadvantaged in the EDFacts data, the school's direct certification rate, the school's Title I status, and the child poverty rate in the school's census tract (as measured by the 5-year ACS) (details on the imputation process are available in Reardon et al, 2024). District FRLE rates are computed as an enrollment-weighted average of the LISD FRLE rates among each district's schools.

There are a number of differences between district poverty as measured by federal free lunch participation and the local area poverty rates on which the Title I program relies. The federal FRLE rates are based on the number of students who apply for free-lunch (or are imputed for CEP schools, where all students are eligible), whereas the Title I formula percent is based on estimated poverty rates within a district's boundaries. FRLE rates include students with incomes less than 185% of the poverty line while Title I eligibility includes only residents below the poverty line. FRPL is based on data reported by public schools; Title I eligibility is based on all children age 5-17 living in the district boundaries, regardless of whether they attend a district-administered public school (as opposed to not being enrolled, being home-schooled, or attending a private school, a non-district-administered charter school, or a school outside the district). Moreover, Title I eligibility is based in part on sample-based poverty rate estimates (from the ACS), whereas FRLE rates are based on administrative counts from the schools.

Thus, we have two measures of local community poverty: the share of students receiving federal subsidized lunches and the local area poverty rates provided by Census. In our analysis, we use both as controls for local area poverty.

Percent of 2020-21 School Year Remote/Hybrid: We created a measure of the share of the 2020-21 school year each district was operating remotely or some hybrid or remote and in-person instruction by combining two data sources. The Return to Learn (R2L) tracker, assembled by the American Enterprise Institute (AEI), includes weekly district-level data on mode of instruction (in-person, hybrid, or remote) from August 2020 through June 2021 for 98 percent of enrollment in U.S. school districts with 3 or more schools. The R2L data are based on public information released by school districts and define a district as remote if no

students older than first grade had an in-person option. We take the average weekly value from the week of September 7, 2020, through the week of June 7, 2021, the period during which 95%+ of included districts have available data.

Our second source of data, the COVID-19 School Data Hub (CSDH) tracks whether a school or district was remote, hybrid, or in-person in 48 states throughout the 2020-21 school year. The CSDH data are based on a survey of state education agencies, and as a result, the CSDH data vary substantially by state in terms of frequency (ranging from weekly to semesterly) and unit (district or school) of available data. The CSDH data define a district as remote if "all or most" students participated in virtual schooling. We take the average value from October 2020 through May 2021, the months in which all states have available data.

Each measure is likely subject to error. Under the assumption that such errors would be independent, we take the average of the R2L and CSDH values to average out the noise in each. We impute the average when either value is missing. For example, if a district is missing R2L values, we regress the R2L/CSDH average value on the CSDH value among districts that have both values and use the prediction from this regression to impute that district's average value.

IV. How the Title I Program Works

Because the federal pandemic relief was based on Title I allocations, the legislative compromises which have shaped the Title I funding formula over the years also determined how the federal relief dollars were distributed. Figure 4 reports the mean Title I allocation per population aged 5-17 by the percentage of school-age children eligible for Title I for each of seven states.[5] Title I allocations per child are typically zero for the lowest poverty districts (those with fewer than 2 percent eligible children). The relationship between eligibility and allocations steepens at five and 15 percent eligible children. The reason is that there are four types of Title I grants (basic, targeted, educational finance incentive grants and concentrated grants), each of which has a different eligibility threshold. A district becomes eligible for a basic grant starting at 2 percent eligible children. The targeted and education finance incentive grants start at 5 percent; the concentrated grants start at 15 percent. As each type of grant is layered in, the relationship between grant dollars and the percentage of children who are eligible becomes steeper.

---

[5]   These figures were fitted with a linear spline function, with knots at 2 percent, 5 percent and 15 percent eligible. Although not a perfect fit (for instance, large districts can qualify for a concentrated grant even if they have fewer than 15 percent poor children, as long as they have more than 6500 eligible children), the fitted spline functions shown in Figure 4 are a good summary. The fitted splines explain 98 percent of the variance in Title I grants per population for those districts not subject to hold harmless provisions.

FIGURE 4: TITLE I ALLOCATIONS PER POPULATION AGED 5-17 BY STATE
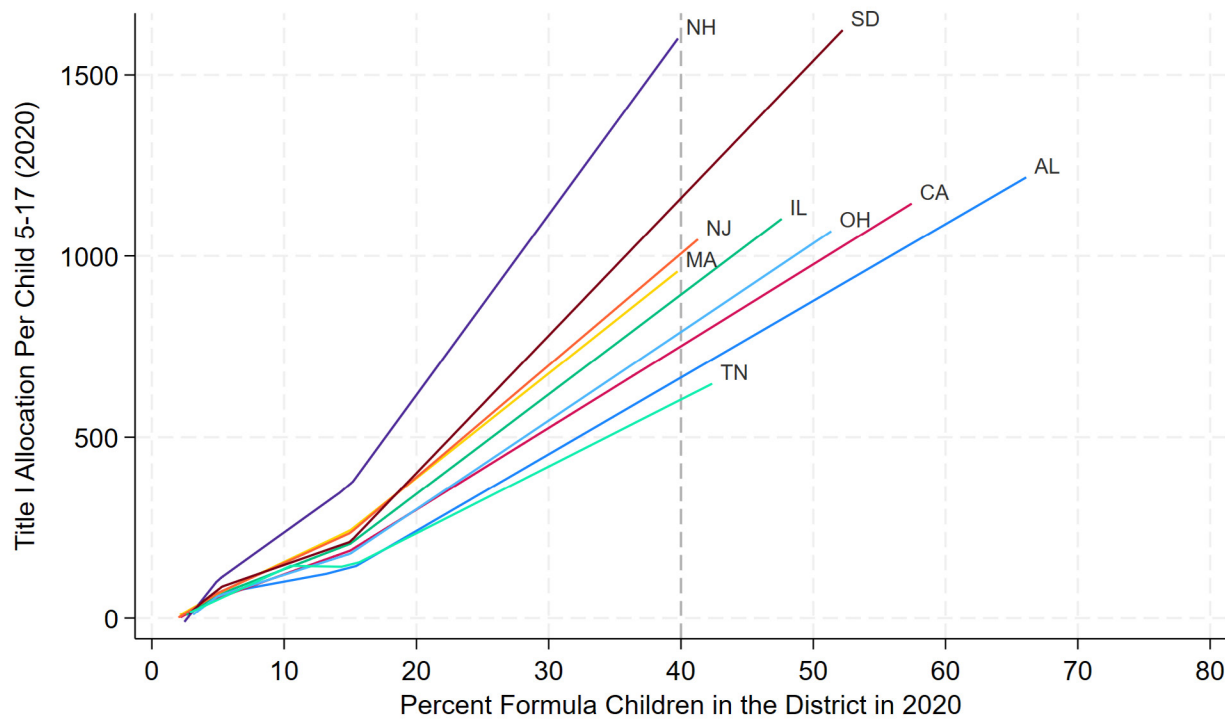
Note: These figures were based on actual Title I allocations in FY 2020, fitted with a linear spline function, with knots at 2 percent, 5 percent and 15 percent eligible children. Although not a perfect fit (for instance, large districts can qualify for a concentrated grant even if they have fewer than 15 percent poor children, as long as they have more than 6500 eligible children), the splines are a good summary, explaining 98 percent of the variance in Title I grants per population for districts not subject to hold harmless provisions.

At any given percentage of eligible children (equivalent to drawing a vertical line in Figure 4), Title I grants vary depending upon the state where the district is located. For instance, a district with 40 percent of children meeting the eligibility formula would have received $572 per child in Tennessee, $652 per child in Alabama, $742 in California, $769 in Ohio, $870 in Illinois, $957 in Massachusetts, $1069 in South Dakota and $1602 in New Hampshire.[6] In other words, for very poor districts, there is roughly a $1000 difference in Title I allocations per child for those in New Hampshire vs. those in Alabama or Tennessee.

Such differences in state formulae are driven by two primary factors: state average per pupil expenditures and the minimum grants for small states.[7] In general, when states increase their average per pupil spending, districts will receive more Title I funding for each poor child. Higher spending per pupil the primary reason why districts in Massachusetts or Illinois receive more funding than districts in Alabama and Tennessee with the same percentage of eligible children.

However, poor districts in small states also benefit tremendously from the way the Title I funding formula

---

[6]   For an excellent explanation of the Title I formula, see Gordon and Reber (2023).

[7]   On paper, there are other factors that matter as well, such as the "state equity factor" (based on the coefficient of variation in expenditures per student across districts in the state) and the "state effort factor" (a function of the ratio of education spending per child and per capita income). However, as Gordon and Reber (2023) show, those adjustments have little effect on the state differences illustrated in Figure 4.

works. The states with the largest Title I grants in Figure 4 (New Hampshire and South Dakota) are not particularly high spending states. One reason is that federal law guarantees that small states receive a minimum share of appropriations for each of the four Title I grant programs. But, in Figure 4, it is not just the average grant that is higher in South Dakota and New Hampshire; the slope is steeper—meaning that poor districts especially benefit from the small state minimum. The reason for the steeper slope is an indirect result of the fact that Title I is not fully funded. For example, based on program rules, districts were eligible for 7 times more funding under the Title I basic grant program than Congress appropriated in FY 2021 (Gordon and Reber 2023). When appropriations fall short of authorized grants, Title I allocations are proportionally reduced within a state. Thus, when a state's appropriation is boosted by the small state minimum, the allocations are increased proportionally, meaning that districts in small states receive allocations closer to authorized formula. The states that benefited from the small state minimums in FY 2020 were Vermont, Wyoming, North Dakota, New Hampshire, DC, Alaska, South Dakota and Montana (Gordon and Reber 2023).

Finally, all four types of Title I grants are subject to "hold harmless" provisions. Depending on its current poverty rate, a district is guaranteed a minimum of between 85 and 95 percent of their Title I grant from the previous year. That means that a positive fluctuation in a district's poverty rate will not only increase a district's grant in that year, but often for several years afterward. The hold harmless provisions also dampen the effect of negative fluctuations.

FIGURE 5. ESSER ALLOCATIONS PER STUDENT BY PERCENT OF STUDENTS RECEIVING SUBSIDIZED LUNCHES
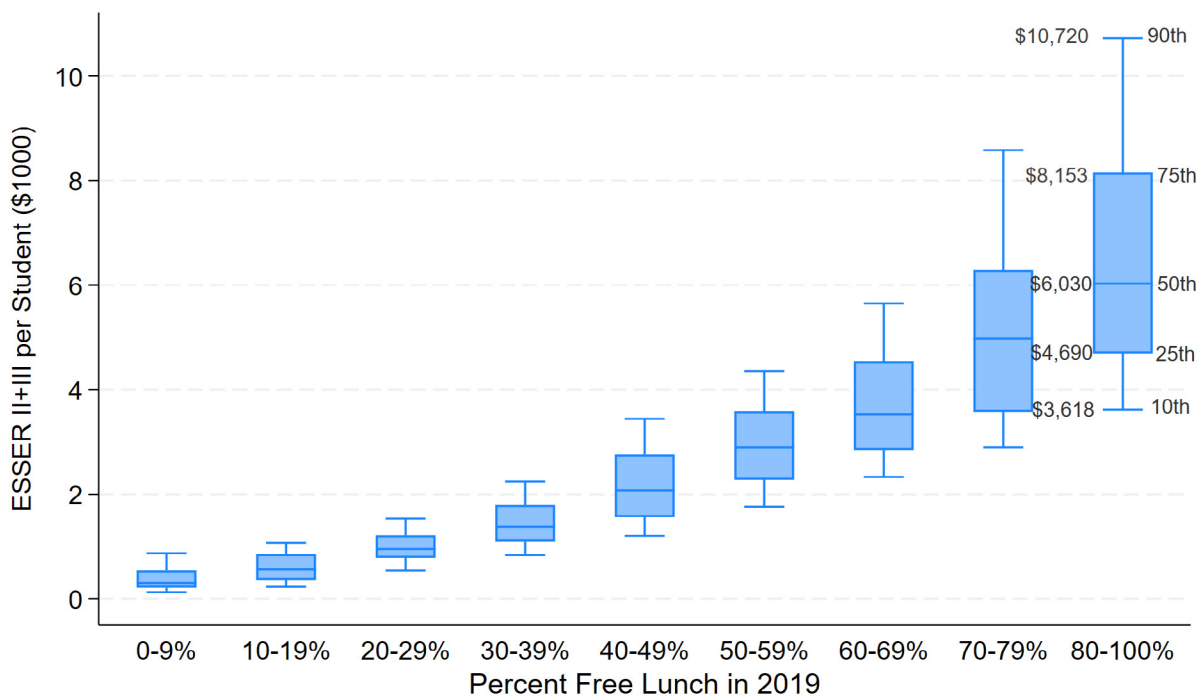
Note: We divide ESSER II and ARP allocations by district total enrollment in 2022 from the Common Core of Data. The percentages of students receiving federal subsidized lunches are from Reardon et al. (2024). The estimates are weighted by district size.

In Figure 5, we report the variation in ESSER II and ARP grants received per student by the percentage of students receiving federally subsidized lunches. Note that while average funding per student is higher in

higher poverty districts, the range in the size of grants per student also fans out. In the highest poverty districts (those with 80-100 percent of students receiving subsidized lunches), the 10th percentile district received $3618 per eligible student. However, the 90th percentile district in that group received over $7,000 more per student ($10,720). An important driver of the fanning out is the difference in state Title I allocations per student illustrated in Figure 4.

Total funding for the Title I program was $16 billion in FY 2020, while the total funding for ESSER II and ARP was $175 billion—slightly more than 10 times more. Because the federal relief funds were distributed proportional to each district's FY 2020 Title I allocation, the relief packages essentially multiplied the differences in the state formulae by 10: a $1000 difference in Title I grant per student became a $10,000 difference in federal pandemic relief per student. In our analysis, we use this variation to investigate the impact of federal pandemic relief on student achievement.

V. Regressing Changes in Achievement on Federal Pandemic Relief, Controlling for District Characteristics

We take two different approaches to measuring the impact of ESSER spending on recovery during 2022-23. In this section, we focus on the relationship between the change in achievement between 2022 and 2023 and federal pandemic relief spending, controlling for district characteristics. In a second set of analyses below, we identify a subset of high-poverty districts with large ESSER grants and compare their recovery to a group of similarly high-poverty districts with much smaller grants, but with similar trends in achievement between 2016 and 2022. In this section, we prioritize controlling for district characteristics. In the next section, we prioritize controlling for underlying trends in achievement. As we will show, we find similar answers using the two approaches.

Suppose student achievement in any year is a function of district expenditure per student in that year (Expend$_{it}$), student characteristics (X$_{it}$) and a district fixed effect (δ_i). The district fixed effect, δ_i, is meant to capture the many unmeasured determinants of achievement in that district which remain fixed over time:

$$(1) \quad S_{it} = \alpha_t + \beta_1 Expend_i + X_{it}\gamma_t + \delta_i + \varepsilon_{it}$$

where the outcome measure, S_{it}, is measured in student-level standard deviations in 2019, i is a subscript for district. Taking the difference in equation (1) between 2022 and 2023, the change in achievement could then be expressed as:

$$(2) \quad S_{i23} - S_{i,22} = \tilde{\alpha} + \beta_1(Expend_{i23} - Expend_{i22}) + X_{i22}\tilde{\gamma} + \tilde{\varepsilon}_i$$

where $\tilde{\alpha}$, $\tilde{\gamma}$ and $\tilde{\varepsilon}_i$ represent the difference in intercepts, coefficient on district characteristics and the error terms between 2023 and 2022 respectively. Note that the district fixed effects no longer play a role. In other words, by focusing on changes in outcomes and increases in expenditures, we are implicitly controlling for many unmeasured determinants of student outcomes, as long as those factors remain fixed over time. Nevertheless, if the role of unmeasured district characteristics were changing between 2022 and 2023, such that unmeasured determinants of the change in achievement ($\tilde{\varepsilon}_i$) was correlated with the change in expenditures, $Expend_{i23} - Expend_{i22}$, our estimate of β_1 could still be biased. (That is a primary reason

we check our results with the second strategy below, finding districts with similar trends in achievement through 2022.)

Because the data on district spending have not been released by the National Center on Education Statistics, we use the ESSER allocation per student and spending per student as proxies for increases in district revenues and expenditures respectively. This is a reasonable assumption in the case of expenditures from state revenues, as maintenance of effort provisions in the federal law required states to maintain spending.[8] If local governments tried to reduce their contribution, it would negatively impact their Title I revenues in the future. Nevertheless, if some local governments did cut back their contributions, a $1000 allocation per student in ESSER dollars would have resulted in less than $1000 in additional expenditures, leading us to understate the impact per federal dollar in aid.

To estimate the impact per dollar spent for the average student, we weight school districts in the regression models by the size of their grade 3-8 enrollment (the grades for whom our test score measures apply). The estimates from these regressions are shown in Tables 2 and 3. As reported in the first column of Table 2, the coefficient on ESSER allocation per student is .0045 SD per $1000 in math.

In the second column, we split the ESSER II+III allocation per student into two parts: the amount spent as of May/June 2023 and the amount not yet spent. In column (2), the coefficient on ESSER dollars spent per student is .0059 SD per $1000. In other words, among districts that received the same allocations, those that spent more during the 2022-23 school year saw faster growth during the 2022-23 school year. The coefficient on dollars unspent serves as a sort of placebo test: if we had found that dollars not yet spent were related to improved achievement, that would suggest there is some unmeasured factor related both to ESSER allocations and improved achievement. We find no association between dollars unspent and achievement gains.

As noted above, our equating of state test results may result in a state-level estimation error, common to all districts in a state. To address the concern that this error might be correlated with between-state differences in spending, we add controls for state fixed effects in column (3). The results are unchanged.

In column (4), we add controls for district characteristics and state fixed effects. The district characteristics include bins for the percentage of students receiving federal subsidized lunch (0-10, 11-20, …, 90-10), the log of enrollment in grades 3-8 (to control for district size), the percent of students who are Black, the percent Hispanic, the urbanicity of the district (in four categories), as well as the changes in log enrollment, percent black, percent Hispanic and percent receiving federally subsidized lunch between 2022 and 2023. The results are unchanged.

In column (5), we add controls for the percent of the 2020-21 school year that students were operating in remote or hybrid instructional mode. As noted above, higher poverty districts received more ESSER funding on average. And they also were remote for longer during the 2020-21 school year. Even though the federal funding was not based on the achievement losses, districts that were remote for longer likely received more aid. If the districts who were remote for longer simply bounced back more, we could be overstating the effect of spending. Our findings suggest that the districts that were remote for more of the 2020-21 school year did bounce back somewhat more—.03 SD for a 100 percent difference in percent remote and .02 SD for percent hybrid. However, the coefficient on ESSER spending per student is unchanged when we add a control

_____

[8]   The maintenance of effort requirements under ESSER did not apply to local governments. However, as the Department of Education indicated in E-14 of ESSER-and-GEER-Use-of-Funds-FAQs-December-7-2022-Update-1.pdf (ed.gov), any district that chose to replace state or local funds with federal ESSER funds risked failing to meet the maintenance of effort requirement under the Title I program itself.

for remote/hybrid learning.

In column (6), we add controls for the percentage of students in the district estimated by the Census to be eligible for Title I. To allow for a flexible functional form, we include dummies with a bin size of 2 percentage points for percentage of students eligible. The estimated coefficient rises slightly to .0069 SD per $1000 spent.

TABLE 2. ESTIMATING EFFECT OF ESSER ALLOCATION/SPENDING ON DISTRICT CHANGES IN MATH ACHIEVEMENT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Math Achievement											
Analysis Methodology											
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	IV	IV	IV
Spending Effects											
ESSER II+III Allocation per Student ($1,000s)	0.0045*** (0.0008)										
ESSER II+III Spent per Student (as-of June 2023) ($1,000s)		0.0059*** (0.0015)	0.0058*** (0.0014)	0.0051*** (0.0014)	0.0054*** (0.0013)	0.0069*** (0.0017)	0.0069*** (0.0017)	0.0086*** (0.0019)	0.0106*** (0.0025)	0.0168*** (0.0045)	-0.0155 (0.0228)
ESSER II+III Unspent per Student ($1,000s)		0.0026 (0.0018)	0.0009 (0.0016)	-0.0000 (0.0018)	-0.0001 (0.0017)	0.0014 (0.0019)	0.0014 (0.0020)	0.0027 (0.0021)			
Remote/Hybrid Recovery											
% Remote in 2020-21 School Year					0.0326*** (0.0086)	0.0306*** (0.0086)	0.0307*** (0.0087)	0.0300*** (0.0088)	0.0304*** (0.0086)	0.0300*** (0.0087)	0.0322*** (0.0096)
% Hybrid in 2020-21 School Year					0.0228*** (0.0062)	0.0216*** (0.0060)	0.0217*** (0.0061)	0.0212*** (0.0061)	0.0209*** (0.0061)	0.0205*** (0.0061)	0.0230*** (0.0066)
Included Covariates											
State Fixed-Effects		X	X	X	X	X	X	X	X	X	X
Log Enrollment, % Black, % Hispanic FRPL Deciles, Rural, Town, Suburb Starting School Year			X	X	X	X	X	X	X	X	X
Change in Log Enrollment, % Black, % Hispanic, % FRPL over period			X	X	X	X	X	X	X	X	X
% Remote/Hybrid				X	X	X	X	X	X	X	X
Formula Percent - Every 2% (2020)						X	X	X	X	X	X
EB Trend 2016 to 2019							X	X	X	X	X
ESSER II+III Allocation Per Student Limit ($1,000's)								16	16	16	16
Instrument									ESSER Allocation	Title I Per Pop 5-17 Using 2020 Elig (No HH)	Title I Residual (2020)
N	5507	5507	5507	5507	5507	5507	5507	5498	5498	5498	5498
R2	0.032	0.033	0.194	0.223	0.229	0.242	0.242	0.242	0.241	0.234	0.187

Note: The dependent variable is change in mean achievement between 2022 and 2023. Observations are weighted by enrollment in grades 3-8 in 2022. Robust standard errors are reported in parentheses.

TABLE 3. ESTIMATING EFFECT OF ESSER ALLOCATION/SPENDING ON DISTRICT CHANGES IN READING ACHIEVEMENT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Reading Achievement											
Analysis Methodology											
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	IV	IV	IV
Spending Effects											
ESSER II+III Allocation per Student ($1,000s)	0.0024										
	(0.0017)										
ESSER II+III Spent per Student (as-of June 2023) ($1,000s)		0.0101**	0.0034**	0.0049***	0.0050***	0.0051***	0.0051***	0.0049***	0.0045*	0.0049	-0.0053
		(0.0042)	(0.0013)	(0.0011)	(0.0011)	(0.0015)	(0.0015)	(0.0019)	(0.0023)	(0.0046)	(0.0288)
ESSER II+III Unspent per Student ($1,000s)		-0.0082**	-0.0022	-0.0003	-0.0004	-0.0006	-0.0006	-0.0007			
		(0.0038)	(0.0014)	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0018)			
Remote/Hybrid Recovery											
% Remote in 2020-21 School Year					0.0141*	0.0116	0.0117	0.0126*	0.0125*	0.0124*	0.0132*
					(0.0076)	(0.0075)	(0.0075)	(0.0076)	(0.0075)	(0.0075)	(0.0079)
% Hybrid in 2020-21 School Year					-0.0028	-0.0024	-0.0023	-0.0021	-0.0020	-0.0020	-0.0010
					(0.0052)	(0.0051)	(0.0051)	(0.0051)	(0.0050)	(0.0050)	(0.0059)
Included Covariates											
State Fixed-Effects		X	X	X	X	X	X	X	X	X	X
Log Enrollment, % Black, % Hispanic FRPL Deciles, Rural, Town, Suburb Starting School Year			X	X	X	X	X	X	X	X	X
Change in Log Enrollment, % Black, % Hispanic, % FRPL over period % Remote/Hybrid				X	X	X	X	X	X	X	X
Formula Percent - Every 2% (2020) EB Trend 2016 to 2019						X	X	X	X	X	X
ESSER II+III Allocation Per Student Limit ($1,000's)								16	16	16	16
Instrument									ESSER Allocation	Title I Per Pop 5-17 Using 2020 Elig (No HH)	Title I Residual (2020)
N	5127	5127	5127	5127	5127	5127	5127	5117	5117	5117	5117
R2	0.008	0.031	0.439	0.458	0.459	0.467	0.467	0.466	0.466	0.466	0.457

Note: The dependent variable is change in mean achievement between 2022 and 2023. Observations are weighted by enrollment in grades 3-8 in 2022. Robust standard errors are reported in parentheses.

In column (7), we control also for the district's trend in achievement prior to 2022. We use the linear trend in achievement between 2016 and 2019. The results are unchanged. Note that we control more rigorously for prior trends in achievement in the synthetic control estimates in the next section.

There are a few districts with unusually high ESSER allocations per student, which have considerable leverage influencing the coefficient estimate. In column (8), when we drop districts receiving more than $16,000 in ESSER funds per student, the coefficient rises to .0086.[9]

In the remaining columns of Table 2, we estimate the impact of ESSER spending using two-stage least squares. With the two-stage least squares analyses, we isolate the effect of differences in federal relief spending driven by three specific sources: total ESSER allocations per student, state Title I funding formulae and seemingly random fluctuations in poverty rate estimates within each district.

In column (9), we instrument for spending using the total allocation of ESSER II and American Rescue Plan dollars per student that each district received. If district-level differences in timing of spending were endogenous—e.g. the districts with the most capable leadership were able to spend a larger share of their funds during the 2022-23 year—then we could be overstating the effect of spending since spending would partially reflect districts' ability to execute. On the other hand, if the districts with the most intractable challenges spent a larger share of their grants during the 2022-23 year, we could be underestimating the payoff. Moreover, there is likely some measurement error in the annual spending estimates, depending on when districts logged individual expenditures. Thus, by using ESSER allocations as an instrumental variable for 2022-23 spending, we isolate the spending differences which were due to the differences in allocations—essentially assuming that each district spent the same share of their federal relief dollars during 2022-23—and adjusting for possible measurement error in the timing of expenditures. The results in column (9) imply that we are not overstating the effect of spending due to endogeneity of timing. Our estimates are slightly higher, implying a gain of .0106 SD per $1000 of spending.

In column (10), we focus on the variation in spending which was due to the differences in state Title I funding formulae. We continue to control flexibly for the percentage of the population eligible for Title I in fiscal year 2020—including dummy variables for each two percentage points. As an instrument, we use the district allocation per population aged 5-17 based on the state formulae, using the share of students eligible for Title I in fiscal year 2020. In doing so, we exclude other sources of variation in spending per student—not only the timing of district spending decisions, but also increases in Title I grants due to hold harmless provisions (which reflect eligibility in prior years) and differences in the ratio of students enrolled in public schools to the number of school-age children estimated by Census to reside in the district (recall that the percentage of eligible children used in Title I is based on the resident population and is not limited to those enrolled in public schools). We estimate that a $1000 difference in spending was associated with a .0168 SD improvement in achievement.

In the last column of Table 2, we instrument for spending using seemingly random fluctuations around a trend in the percentage of children who are eligible. The point estimate based on such variation is not statistically significant, but the standard error is quite large at .023 standard deviations per student.

Table 3 reports similar estimates for reading. The OLS estimates follow a similar pattern to those in math, although are somewhat smaller at .0049 standard deviations per $1000 expenditure in column (8).

---

[9]   We tested for declining payoffs to higher grants using a quadratic function of ESSER spending per student. We could not reject a linear relationship between spending and achievement.

VI. Matching on Prior Achievement Trends (Synthetic Control Group Estimates)

The analysis in the prior section prioritized controlling for district's characteristics. In this section, we prioritize controlling for prior trends in achievement. Specifically, we use synthetic control methods to compare districts with large and small grants, but which had similar prior achievement trends. As illustrated in Figure 5, federal relief varied dramatically—particularly among high poverty districts. Of the 779 high-poverty districts, with more than 70 percent of students receiving free/reduced price lunches, we identified 149 districts with ESSER grants in the top quartile (more than $8,188 dollars per student in ESSER allocation.) We label these the "treated" districts, since they received the largest grants. For comparison, we identified 244 donor districts which received less than $4,563 per student (the bottom quartile of districts when weighted by enrollment).

For each of the 149 treated districts, we created a synthetic control district following Abadie, Diamond and Hainmueler (2010). The synthetic control is a weighted average of donor districts that matches the pre-treatment trend of the treated district. The weights are constrained to be non-negative and are chosen to maximize the match between each treatment district and its comparisons in terms of their pre-treatment achievement. To capture prior trends in achievement, we matched on three measures: the pre-pandemic change in test scores from 2016 to 2019; the pre-pandemic level of test scores in 2019, and the pandemic change in test scores from 2019 to 2022. In addition, we matched on the log of enrollment in grades 3-8 and the proportion of the 2020-21 school year that the district was remote. Finally, we apply an additional bias correction (Abadie and L'hour, 2021) to districts' scores that adjust for remaining discrepancies between the treated and synthetic control districts' characteristics, analogous to the use of regression adjustment after propensity score matching.

For each of the 149 treated districts, we constructed the difference between the treated district and the synthetic control district for two outcomes: their test scores in 2023 and their ESSER spending in 2022-23. We averaged these differences across the 149 treated districts weighting by the enrollment in the treated district to obtain overall estimates of the difference in spending and test scores between treated districts with high ESSER spending and their synthetic control districts with low ESSER spending. The ratio of the test score difference to the spending difference yields an estimate of the impact of spending on test scores (analogous to a Wald estimator) among this set of high poverty districts. Since this estimate is simply a difference in (weighted) means between the treatment districts and donor districts, we calculated standard errors using standard methods that are robust to heteroskedasticity.[10]

---

[10]   More specifically, to generate standard errors, we estimated the following regression with robust standard errors using our full sample of 149 treated and 244 donor districts:

$$S_{i2023}^{bc} = \beta_o + \beta_1 Treat_i + \varepsilon_i$$

where the dependent variable is the bias-corrected test score of each district in 2023 and Treat=1 if the district was a treated district. We weight each treated district (i) by its enrollment in grades 3-8 ($enroll_i$), and weight each donor district (j) by $\sum_i^K w_{ij} enroll_i$, where w_{ij} are the synthetic weights for comparison group district j used for treatment group district i. The estimated impact from this regression ($\hat{\beta}_1$) is equal to the average difference in the outcome between the treated districts and their synthetic control group, weighted by treated district enrollment size. The standard errors from this regression treat the synthetic control weights as known constants and ignore sampling variation. In addition to the change in achievement, we use

Table 4 reports the mean characteristics of the treatment group and the synthetic comparison groups for math. As reported in column (1), the treatment group had mean achievement in math of -.531 (half a standard deviation below the 2019 average). Before applying the weights, the mean achievement of the donor districts was somewhat higher, -.344 SD below the national average in 2019. However, after applying the weights, the 2019 mean achievement for the synthetic controls was much closer to the treatment group, -.522 vs. -531. Likewise, the mean loss in achievement between 2019 and 2022 was -.146 SD in the treatment group and -.141 SD in the synthetic control group after applying the weights. In other words, the weights resulted in combinations of the donor districts which were much more similar in terms of prior trends to the treatment districts.

Although the synthetic control group was similar in prior outcomes, the weighting left substantial differences in other characteristics. For instance, the treatment districts remained quite a bit larger than the synthetic control districts, with mean log enrollment of 8.92 vs. 10.98. The synthetic control group spent an average of 50 percent of the 2020-21 year remote vs. 61 percent for the treatment group. The bias correction method adjusts for these remaining discrepancies.

In Table 5, we report mean characteristics for treatment and comparison groups when the weights are chosen to match on prior reading scores. The treated districts had an average achievement .433 standard deviations below the national average in 2019, had a small loss in achievement of .008 standard deviations between 2016 and 2019 and suffered a loss of .070 standard deviations between 2019 and 2022. The synthetic control districts had similar prior outcomes: mean achievement .425 standard deviations below the national average in 2019, little loss of .008 SD between 2016 and 2019 and a loss of .070 standard deviations between 2019 and 2022.

In Table 6, we report the estimated impacts on the treated districts relative to their controls: in math, we estimated that the treated districts grew by .056 standard deviations more than their comparisons (a statistically significant difference given the standard error of .0193). We also estimated that the treated districts spent $2816 more per student between 2022 and 2023 out of the ESSER dollars. Dividing the .0558 estimated impact on achievement by the $2816 difference in spending yields an implied impact per dollar spent of .0198 SD per $1000 spending per student. In reading, we estimated a similar impact of .0189 SD per $1000 spent per student. Although the estimated impact on reading was statistically significant at the .10 level, it was not significant at the conventional .05 level.[11]

the above framework to estimate differences in allocation per student and spending per student relative to the comparisons.
[11]   We did two additional robustness tests. First, we formed synthetic comparisons for the comparison districts and estimated "impacts" for them. As expected, we found no difference in outcomes. Second, we repeated the exercise with "low-poverty" districts, those with fewer than 20 percent of students receiving federal lunch subsidies. The estimated difference in allocation per student between "treated" districts (those in the top quartile of grants per student) and their synthetic comparisons made up of those in the bottom quartile was much smaller than the high poverty districts, $977 as opposed to $6189 dollars per student. The difference in spending was even smaller: $656 per student. Accordingly, we would expect smaller impacts. And, indeed, we found no statistically significant difference in 2022 to 2023 growth using the methods described above.

TABLE 4. MATH TREATMENT VS. SYNTHETIC CONTROL COMPARISONS

Math Synthetic Control Summary Statistics

	Treatment		Control		Control (Weighted)	
	Mean	SD	Mean	SD	Mean	SD
Primary Outcomes:						
2019 Achievement in Math	-0.531	0.214	-0.344	0.249	-0.522	0.325
2019-22 Change in Math	-0.146	0.099	-0.162	0.109	-0.141	0.108
2016-19 Change in Math	0.006	0.034	0.000	0.033	0.010	0.041
Secondary Outcomes:						
2019 Achievement in Reading	-0.448	0.208	-0.264	0.222	-0.416	0.301
2019-22 Change in Reading	-0.057	0.089	-0.109	0.102	-0.096	0.104
2016-19 Change in Reading	-0.008	0.037	-0.006	0.035	0.002	0.044
Other District Characteristics:						
Log Total Enrollment (2022)	10.98	1.97	9.10	1.01	8.92	1.47
% Black (2022)	32.9%	28.0%	10.6%	17.0%	23.5%	33.7%
% Hispanic (2022)	45.8%	29.7%	62.8%	28.2%	55.8%	35.2%
% Free Lunch (2022)	78.6%	6.4%	76.6%	6.5%	76.6%	7.0%
Percent of Population Aged 5-17 Who Met Title I Formula (2020)	31.7%	7.8%	21.2%	4.6%	23.3%	6.5%
Percent of 2020-21 SY in Remote Instruction	61.1%	25.9%	53.9%	31.6%	49.9%	30.6%
Percent of 2020-21 SY in Hybrid Instruction	29.7%	17.3%	30.9%	22.7%	27.6%	15.8%
ESSER II/ARP Dollars Allocated Per Student	$9,834	$1,913	$3,477	$664	$3,645	$695
ESSER II/ARP Dollars Spent Per Student	$5,249	$1,522	$2,103	$673	$2,433	$843
Number of Districts	127		210			
Number of Students	1,151,454		1,036,401			

Note: The sample was limited to districts with more than 70 percent of students receiving federally subsidized lunches. The treatment districts received ESSER II+ARP allocations per student in the top quartile (more than $8188 per student); the synthetic controls were weighted averages of those with ESSER II+ARP allocations in the bottom quartile (less than $4563 per student). The weights were chosen to match the pre-2023 achievement for each treated district.

TABLE 5. READING SYNTHETIC CONTROL GROUP COMPARISONS

Reading Synthetic Control Summary Statistics

	Treatment		Control		Control (Weighted)	
	Mean	SD	Mean	SD	Mean	SD
Primary Outcomes						
2019 Achievement in Reading	-0.433	0.211	-0.258	0.217	-0.425	0.311
2019-22 Change in Reading	-0.070	0.101	-0.117	0.105	-0.070	0.121
2016-19 Change in Reading	-0.008	0.039	-0.008	0.035	-0.008	0.050
Secondary Outcomes						
2019 Achievement in Math	-0.510	0.214	-0.328	0.260	-0.495	0.327
2019-22 Change in Math	-0.148	0.106	-0.170	0.120	-0.116	0.099
2016-19 Change in Math	0.004	0.035	-0.002	0.035	0.007	0.041
Other District Characteristics:						
Log Total Enrollment (2022)	10.97	2.00	9.15	0.95	9.01	1.20
% Black (2022)	30.7%	28.1%	12.2%	18.2%	20.0%	31.2%
% Hispanic (2022)	48.6%	30.0%	62.4%	28.8%	58.4%	31.6%
% Free Lunch (2022)	78.1%	6.3%	77.3%	6.8%	78.7%	6.7%
Percent of Population Aged 5-17 Who Met Title I Formula (2020)	31.0%	7.7%	21.2%	4.4%	22.5%	7.8%
Percent of 2020-21 SY in Remote Instruction	60.2%	25.7%	56.7%	29.0%	53.4%	28.7%
Percent of 2020-21 SY in Hybrid Instruction	30.7%	18.0%	33.0%	23.1%	35.8%	24.0%
ESSER II/ARP Dollars Allocated Per Student	$9,307	$1,326	$3,521	$667	$3,688	$560
ESSER II/ARP Dollars Spent Per Student	$4,994	$1,447	$2,079	$649	$2,509	$607
Number of Districts	108		179			
Number of Students	1,032,665		1,032,326			

Note: The sample was limited to districts with more than 70 percent of students receiving federally subsidized lunches. The treatment districts received ESSER II+ARP allocations per student in the top quartile (more than $8188 per student); the synthetic controls were weighted averages of those with ESSER II+ARP allocations in the bottom quartile (less than $4563 per student). The weights were chosen to match the pre-2023 achievement for each treated district.

Figure 6 illustrates the synthetic control results by plotting the trend over time in the average bias corrected test scores for the treatment districts and the synthetic control districts. In both math and reading, the treatment and control districts followed a similar trend between 2016 and 2022 (by construction) but then diverge between 2022 and 2023, with both math and reading test scores growing by about .05 s.d more in treatment districts with high ESSER spending compared to control districts with lower ESSER spending.

TABLE 6: TREATED VS. SYNTHETIC COMPARISON DISTRICTS

	Math	Reading
Impact on 2023 Achievement	0.0558	0.0471
	(0.0193)	(0.0263)
Spending per Student	$2816	$2486
	($396)	($251)
Implied Impact per $1000	0.0198	0.0189
	(0.0060)	(0.0116)
N	337	287

Note: The above are differences between treated and synthetic comparison districts, matched on achievement in 2019, and changes in mean achievement between 2016-19 and 2019-22. All sample districts had more than 70 percent of students receiving federal subsidized lunches. The treated districts were in the top quartile of federal ESSER II+ ARP allocations per student (more than $8188) and the comparisons were in the bottom quartile (receiving less than $4563 per student). The estimates were bias corrected based on district size and share of the 2020-21 year that districts were remote (Abadie and L'Hour 2021).

FIGURE 6: HIGH POVERTY/HIGH ESSER DISTRICTS VS. HIGH POVERTY/ LOW ESSER SYNTHETIC CONTROLS
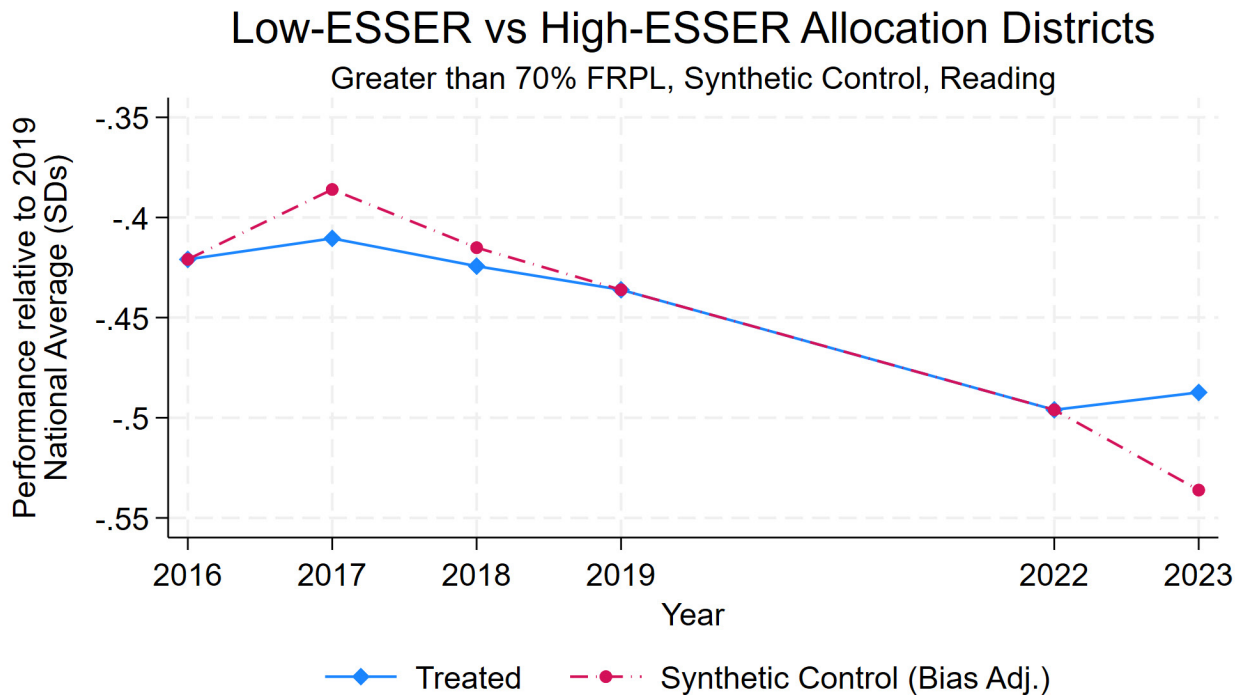
Low-ESSER vs High-ESSER Allocation Districts
Greater than 70% FRPL, Synthetic Control, Reading

Note: Treatment group is the top quartile of ESSER allocation among districs with >70% FRPL. Donor group for synthetic control is the bottom quartile of ESSER allocation among districts with >70% FRPL. Donor group is weighted using average OL in 2016-2019, OL change from 2019-2022, % remote, and log enrollment in grades 3-8. All districts weighted by grade 3-8 enrollment in 2022.

Low-ESSER vs High-ESSER Allocation Districts
Greater than 70% FRPL, Synthetic Control, Math
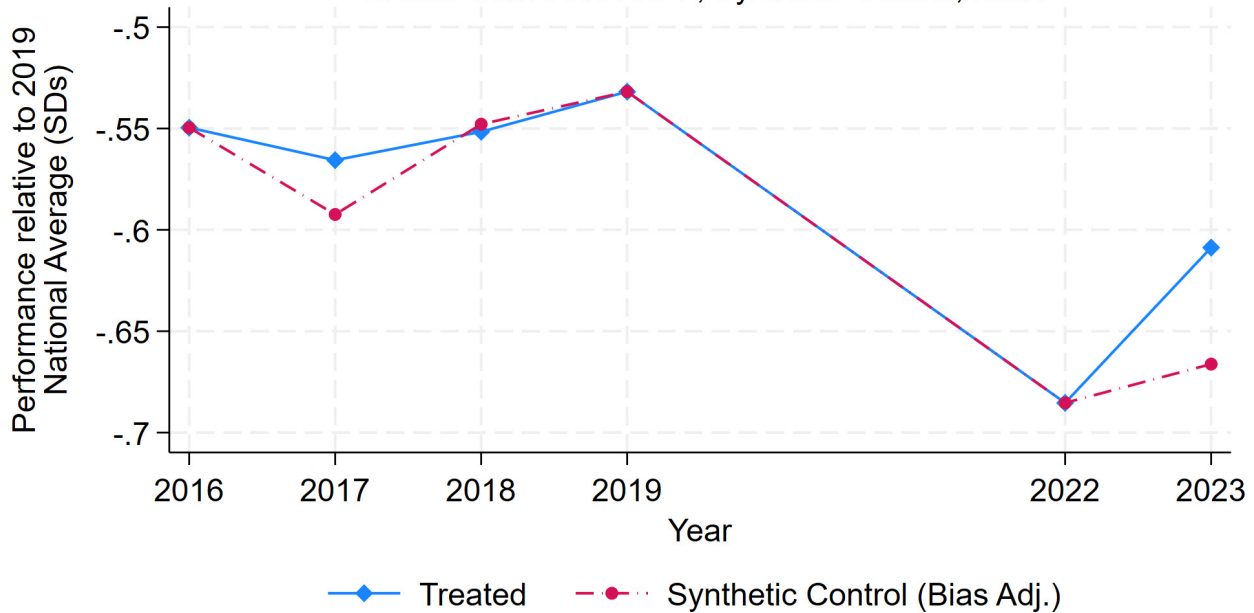
Note: Treatment group is the top quartile of ESSER allocation among districs with >70% FRPL. Donor group for synthetic control is the bottom quartile of ESSER allocation among districts with >70% FRPL. Donor group is weighted using average OL in 2016-2019, OL change from 2019-2022, % remote, and log enrollment in grades 3-8. All districts weighted by grade 3-8 enrollment in 2022.

VII. Comparison to Prior Literature

In Figure 7, we plot the confidence intervals for each of our point estimates relative to the meta-analytic mean of .0079 in Jackson and Mackevicius (2024). In most cases, we can reject the hypothesis that the federal pandemic relief dollars had no impact on achievement (zero is not included in the confidence interval). The only two exceptions are the estimates of impact on reading achievement from the instrumental variable estimates and the synthetic control groups. But, in all cases, we cannot reject the hypothesis that the effect of federal pandemic relief had the same impact per student as the mean impact identified by Jackson and Mackevicius' literature review.

FIGURE 7. COMPARING POINT ESTIMATES TO JACKSON AND MACKEVICIUS

Impact Estimates of ESSER Spending
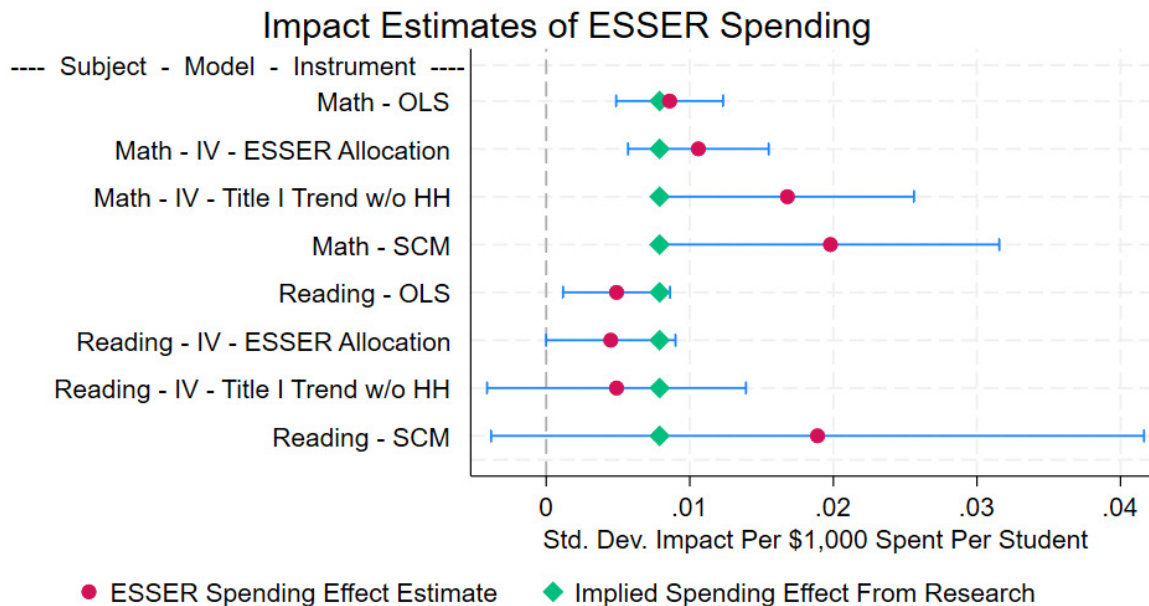
Note: Implied effect from research is derived from 0.0316 standard deviations increase in achievement per $1,000 per student over 4 years per the findings in Jackson and Mackevicius (2024).

In their meta-analysis, Jackson and Mackevicius include studies involving capital spending, finding an effect similar to current spending when amortized over the life of the project. However, they assumed zero impact of capital spending during the first two years to allow for construction delays. In their analysis of district spending plans for the American Rescue Plan dollars, Brooks and Springer (2024) estimated that districts were planning to spend 27 percent on facilities improvements (including HVAC). If we were to adjust our point estimates for the 27 percent capital spending (by dividing by .73), our estimated impacts would be even higher: the .0086 per $1000 estimate in math would become .0117, and the .0048 in reading would become .0066. It could be that the capital spending during 2022-23 and earlier will have follow-on effects on district achievement in 2024 and 2025. Thus, future analyses of the effect of pandemic relief in later years should account for prior capital spending.

VIII. Conclusion

Over the past three years, there have been multiple reports in the press of districts using federal relief dollars for seemingly unintended purposes such as athletic fields[12] (e.g. Associated Press 2021). Some of us have co-authored papers describing implementation challenges and disappointing results from specific catch-up efforts. (e.g. Carbonari et al. 2022). Thus, many are likely wondering whether the ESSER aid truly helped students recover. Our results suggest the spending did have a positive impact on achievement. Indeed, the estimated impact is in line with the prior research on the effect of increased education spending.

However, that finding begs three related questions: How much of the recovery between 2022 and 2023 could

---

[12]   For instance, the organization, Parents Defending Education, posted the top 10 most wasteful ESSER expenditures on its website: https://defendinged.org/investigations/wasteful-esser-expenditures/

be explained by the ESSER spending? And, at current rates of impact per dollar spent, how much would a full recovery cost? And, finally, if additional federal or state spending is needed to complete the recovery, how might it be structured to increase impact per dollar spent?

The average U.S. student lost .149 standard deviations in math achievement between 2019 and 2022. Returning to 2019 levels on the basis of federal pandemic relief alone would require $18,800 per student (dividing the .149 standard deviation loss by .0079 SD per $1000 from Jackson and Mackevicius). When aggregated across the 48 million students who were enrolled in public schools in the U.S., a fully federally funded recovery at that rate would have costed $904 billion—about 5 times more than the $190 billion provided.[13]

But our results imply that would be an overestimate. The average recovery by 2023 is larger than that implied by our estimate of the effect of ESSER spending. There must be other factors—such as parental investments at home, teacher or student effort, perhaps even increases in local spending—which are contributing to the recovery. In Figure 8, we portray the average improvement between 2022 to 2023 for districts in each of ten bins, organized by the percentage of students receiving federally subsidized lunches. With the green dashed line, we report the increase we would have expected based on the federal dollars spent during 2022-23 (multiplying the average ESSER spending per student during 2022-23 by .0086 standard deviations per $1000 spent, the coefficient on ESSER spending in column 8 of Table 2). While the federal relief can explain between one-third to one-half of the improvement in districts with more than 70 percent of students receiving federal lunch subsidies, it explains little to none of improvement in higher income districts, because they did not receive substantial amounts of federal relief. The additional improvement that occurred over and above the estimated effect of spending is portrayed by the gray shaded area in Figure 8. When averaged across all the groups, the additional growth from sources other than federal spending was .03 standard deviations.

Does a .0086 standard deviation improvement in achievement per $1000 expenditure a worthwhile investment for society? Research on the relationship between test scores and earnings suggests that it is. For instance, Neal and Johnson (1996) find that a standard deviation in the AFQT test was associated with roughly a 20 percent difference in earnings at age 26-29 for both men and women. Murnane et al. (2000) find that a standard deviation in 10th grade math scores was associated with a 12 percentage point difference in earnings at age 31. More recently, Watts (2020) finds that a 1 standard deviation in achievement is associated with a 12 percent difference in earnings for mean and women between the ages of 33 and 50. Discounting future earnings back to their current age, Doty et al. (2022) estimated an average present value of lifetime earnings for K-12 students of $1.2 million. Assuming a 12 percent boost in lifetime earnings per standard deviation in achievement, a .0086 standard deviation increase in achievement would be worth $1,238— somewhat more than $1000, although not dramatically so. Other benefits of increased achievement—such as lower arrests, lower teen motherhood (as reported by Doty et al. 2022)—would enhance the social return.

Looking forward, what do our estimates imply about the magnitude of recovery during the 2023-24 school year? A provisional answer to that question requires several assumptions. We start by assuming that districts will have spent their remaining federal funds between spring 2023 and spring 2024 (they are required to obligate it all or return the remainder by September 2024, just a few months later). Moreover, we assume that the effect of spending during the 2023-2024 school year is equal to our estimate for the 2022-23 school year. Under these assumptions, the orange line in Figure 8 portrays the expected additional impact of the 2023-24 spending. When added to prior recovery, our prediction would imply that the federal relief aid will

---

[13]   Shores and Steinberg (2022) also used the Jackson and Mackevicius (2024) meta-analysis to predict that the recovery would cost $930 billion.

have had larger impacts on higher poverty districts and will have helped to narrow the gap which opened up during the pandemic. However, the average highest poverty district, with 80 percent or more students eligible for free/reduced price lunch, is forecast to remain .1 SD (approximately a third of a grade equivalent) behind 2019 levels of achievement in Spring 2024.

Of course, these projections are based on relatively strong assumptions. We do not know how districts spent their remaining ESSER funds in 2023-2024 or if they followed the same strategies they used during the 2022-23 year. The full effect of the ESSER spending will not be clear until we have data on 2023-24 spending patterns and student achievement in Spring 2024.

Like Jackson and Mackevicius (2024), we conclude that many districts spent the pandemic relief dollars in ways which boosted student achievement. But that is different from saying that the dollars had as much impact as they could have had. As noted above, researchers such as Harris (2009) and Guryan et al. (2023) have found higher effectiveness-cost ratios for interventions such as K-3 class size, summer learning and high-dosage tutoring programs. While districts did invest in all three, they also spent the funds on other activities (including worthwhile efforts such as masks and ventilation). Rather than provide general use funds, as with ESSER, future state or federal aid might boost achievement even more by incentivizing districts to invest specifically in evidence-based academic catch-up efforts with higher cost effectiveness, such as extending the school year or summer learning (as Texas has done) or expanding tutoring programs (as Maryland and Virginia have done.)

FIGURE 8. ESTIMATED EFFECT OF FEDERAL AID ON ACADEMIC RECOVERY IN MATH 2022-23
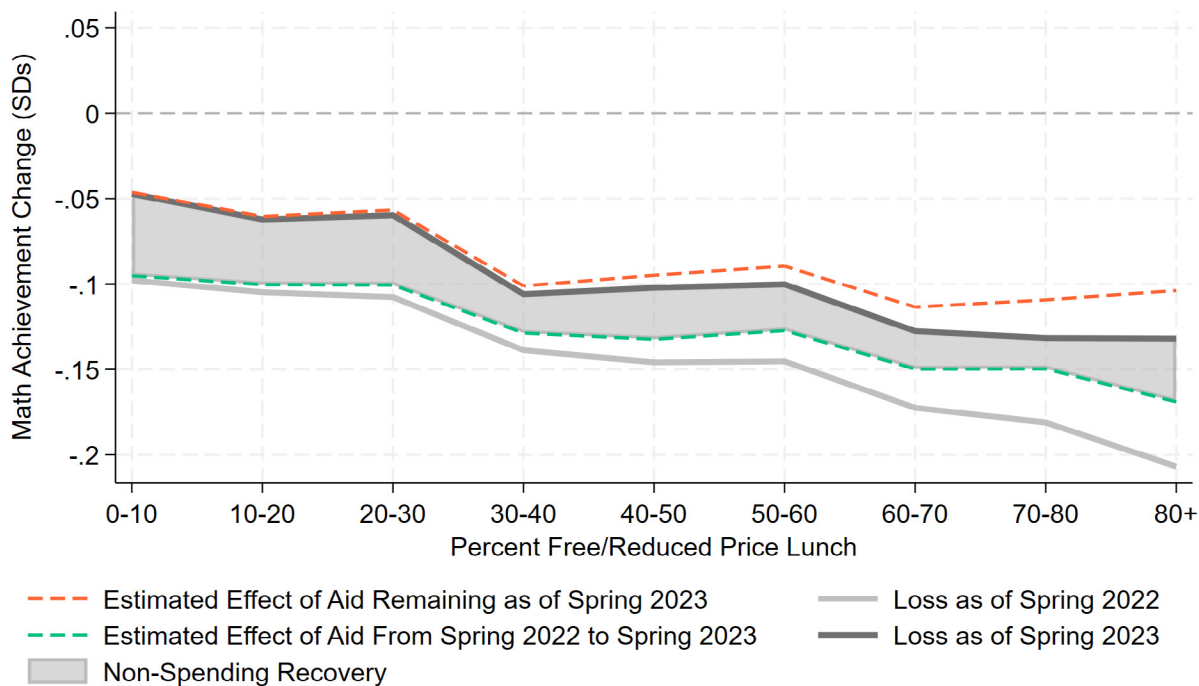
Note: The grey lines above portray our estimates of the actual loss between 2019 and 2022 and the loss remaining as of Spring 2023. The green line is derived by multiplying dollars spent per student between 2022 and 2023 by .0086, our estimated effect of ESSER spending on recovery, and adding this amount to the loss as of Spring 2022—the difference between the green and light grey line is the portion of the academic recovery attributable to federal aid. The orange

line is a forecast, based on the assumption that districts spent all their remaining aid between Spring 2023 and Spring 2024 and spent it in ways that were equally effective to our estimates of earlier spending. The shaded area is the improvement between 2022 and 2023 in excess of that predicted by federal spending alone.

References

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

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects, *Journal of Business and Economic Statistics*, 29 (1) :1-11.

Abadie, A., and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data, *Journal of the American Statistical Association*, 116 (536), 1817–1834.

Associated Press (2021). Schools use federal pandemic relief funds to pay for athletics projects October 6, 2021. As published in Washington Post here.

Bell, W. R., Basel, W. W., & Maples, J. (2016). An overview of the U.S. Census Bureau's small area income and poverty estimates program. In Monica Pratesi (ed.) Analysis of Poverty Data by Small Area Estimation.Wiley & Sons.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of success for all. *American Educational Research Journal,* 44(3), 701-731. https://doi.org/10.3102/0002831207306743

Brooks, C. D., & Springer, M. G. (2024). ESSER-ting preferences: examining school district preferences for using federal pandemic relief fundings. (EdWorkingPaper: 24-913). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/mpm0-1a97

Carbonari, M. V., Davison, M., DeArmond, M.,Dewey, D., Dizon-Ross, E., Goldhaber, D., Hashim, A., Kane, T. J., McEachin, A., Morton, E., Patterson, T., Staiger, D. O. (2022). The challenges of implementing academic COVID recovery interventions: Evidence from the Road to Recovery Project. CALDER Working Paper No. 275-1222

Coleman, J. S., Campbell, E.Q., Hobson, C. J., McPartland, J., Mood, A.M., Weinfeld, F. D., & York, R.L. (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.

Doty, E., Kane, T. J., Patterson, T., & Staiger, D. O. (2022). What do changes in state test scores imply for later life outcomes? NBER Working Paper No. 30701, National Bureau of Economic Research, Cambridge, MA December.

Fahle, E., Kane, T. J., Patterson, T., Reardon, S., Staiger, D.O., & Stuart, E. (2023). School district and community fgactors associated with learning loss during the COVID-19 pandemic. Center for Education Policy Research Brief, Harvard University, May 2023.

Fahle, E., Kane, T.J., Reardon, S. & Staiger, D. O. (2024a). The first year of pandemic recovery: A district-level analysis. Center for Education Policy Research Brief, Harvard University, January 2024.

Fahle, E., Reardon, S., Shear, B., Ho, A., Min, J., & Kalogrides, D. (2024b). "Stanford Education Data Archive Technical Documentation, SEDA2023." Available at https://edopportunity.org/docs/seda2023_documentation_20240130.pdf.

Gordon, N. & Reber, S. (2023). Title I of ESEA: How the formulas work. All4Ed. https://all4ed.org/publication/title-i-of-esea-how-the-formulas-work/

Guryan, J., Ludwig, J., Bhatt, M.P., Cook, P.J., Davis J.M., Dodge, K., Farkas, G., Fryer Jr., R.G., Mayer, S., Pollack, H., Steinberg, L., and Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents.

American Economic Review , 113 (3): 738–765. https://doi.org/10.1257/aer.20210434

Harris, D. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, 31 (1): 3-29.

Handel, D. V., & Hanushek, E. A. (2023). U.S. school finance: Resources and outcomes. In *Handbook of the Economics of Education*. Volume 7, edited by Eric A Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland: 143-226.

Hodrick, R. J. & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. J*ournal of Money, Credit and Banking*, 29 (1) : 1-16.

Jackson, C. K. & Mackevicius, C. (2024). What impacts can we expect from school spending policy? Evidence from evaluations in the United States. *American Economic Journal: Applied Economics*, 16 (1): 412–446. https://doi.org/10.1257/app.20220279

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19 (4): 547–568. https://doi.org/10.1002/1520-6688(200023)19:4%3C547::AID-PAM2%3E3.0.CO;2-%23

National Assessment of Educational Progress (NAEP), 1990-2022 Mathematics Assessments.

Neal, D. A. & Johnson, W.R. (1996). The role of premarket factors in Black-White wage differences. *Journal of Political Economy*, 104 (5): 869-895.

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. NBER Working Paper 27476. https://www.nber.org/papers/w27476.

Reardon, S.F., Shear, B.R., Castellano, K.E., & Ho, A.D. (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics*.

Reardon, S.F., Kalogrides, D., & Ho. A.D. (2021). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Journal of Educational and Behavioral Statistics*, 46(2): 135-137. https://doi.org/10.3102/1076998619874089.

Reardon, S. F., Owens, A., Kalogrides, D., Jang, H, & Tom, T. (2024). The longitudinal imputed school dataset (LISD), Version 1.0. Data available at https://edopportunity.org/segregation/data/. Documentation at https://stacks.stanford.edu/file/druid:gm391gj1253/LISD_geo_crosswalk_documentation_1.0.pdf.

Shear, B.R., & Reardon, S.F. (2021). Using pooled heteroskedastic ordered probit models to improve small-sample estimates. *Journal of Educational and Behavioral Statistics*, 46(1):3-33. https://doi.org/10.3102/1076998620922919.

Shores, K & Steinberg, M P. (2022). Fiscal dederalism and K–12 education funding: Policy lessons from two educational crises. *Educational Researcher*, 20 (10): 1–8. DOI: 10.3102/0013189X221125764.

Snyder, T., Dinkes, R., Sonnenberg, W., and Cornman, S. (2018). Study of the Title I, Part A Grant Program Mathematical Formulas (2019-2016). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Watts, T. (2020). Academic achievement and economic attainment: Reexamining associations between test scores and long-run earnings. *AERA Open* 6 (2):1-16. https://doi.org/10.1177/2332858420928985.

EDUCATION RECOVERY
SCORECARD