February, 2016

The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms

C. Kirabo Jackson, *Northwestern University*

THE EFFECTS OF SCHOOL SPENDING ON EDUCATIONAL AND ECONOMIC OUTCOMES: EVIDENCE FROM SCHOOL FINANCE REFORMS*

C. Kirabo Jackson,
Rucker C. Johnson,
Claudia Persico

Since the Coleman Report, many have questioned whether public school spending affects student outcomes. The school finance reforms that began in the early 1970s and accelerated in the 1980s caused dramatic changes to the structure of K–12 education spending in the United States. To study the effect of these school finance reform–induced changes in public school spending on long-run adult outcomes, we link school spending and school finance reform data to detailed, nationally representative data on children born between 1955 and 1985 and followed through 2011. We use the timing of the passage of court-mandated reforms and their associated type of funding formula change as exogenous shifters of school spending, and we compare the adult outcomes of cohorts that were differentially exposed to school finance reforms, depending on place and year of birth. Event study and instrumental variable models reveal that a 10% increase in per pupil spending each year for all 12 years of public school leads to 0.31 more completed years of education, about 7% higher wages, and a 3.2 percentage point reduction in the annual incidence of adult poverty; effects are much more pronounced for children from low-income families. Exogenous spending increases were associated with notable improvements in measured school inputs, including reductions in student-to-teacher ratios, increases in teacher salaries, and longer school years. *JEL* Codes: J10, I20, H7.

I. Introduction

Public K–12 education is one of the largest single components of government spending (OECD 2013), and differences in school's financial resources across neighborhoods are often cited as key contributors to achievement gaps by parental socioeconomic status and race/ethnicity. However, since the Coleman Report (Coleman et al. 1966), researchers have questioned whether

*We thank the PSID staff for access to the confidential restricted-use PSID geocode data, and confidential data provided by the National Center for Education Statistics, U.S. Department of Education. This research was supported by research grants received from the National Science Foundation under Award Number 1324778 (Jackson), and from the Russell Sage Foundation (Johnson). We are grateful for helpful comments received from Larry Katz, David Card, Caroline Hoxby, Jon Guryan, Diane Schanzenbach, several anonymous referees, and seminar participants at UC Berkeley, Harvard University, NBER Summer Institute, and Institute for Research on Poverty Summer Workshop.

The Quarterly Journal of Economics (2016), 157–218. doi:10.1093/qje/qjv036.
Advance Access publication on October 1, 2015.

increased school spending improves student outcomes. The report employed data from a cross-section of students in 1965–1966 and showed that variation in per pupil spending was unrelated to variation in student achievement on standardized tests. Since then, how school spending affects student academic performance has been extensively studied. Hanushek (2003) reviews the more recent literature published on this question, and his conclusions echo those of Coleman et al. (1966).

We present fresh evidence on the enduring question of whether, how, and why school spending affects student outcomes. We focus our analysis on the effects of public school spending. The objectives of this paper are threefold: we aim to (i) isolate exogenous changes in school district per pupil spending that are unrelated to unobserved determinants of student outcomes, (ii) document the relationship between these exogenous changes in spending and the adult outcomes of affected children, and (iii) shed light on mechanisms by documenting the changes in observable school inputs through which any public school spending effects might emerge.

Given that adequate school funding is a necessary condition for providing a quality education, the lack of an observed positive relationship between school spending and student outcomes is surprising.[1] However, there are two key attributes of previous national studies that might limit the ability to draw firm conclusions from their results. The first limitation is that test scores are imperfect measures of learning and may be weakly linked to adult earnings and success in life. Indeed, recent studies have documented that effects on long-run outcomes may go undetected by test scores (e.g., Ludwig and Miller 2007; Deming 2009; Jackson 2012; Chetty et al. 2011; Heckman, Pinto, and Savelyev 2014). We address the limitations of focusing on test scores as our main outcome by looking at the effect of school spending on long-run outcomes such as educational attainment and earnings.

The second limitation of previous work is that most national studies correlate actual changes in school spending with changes

1. Potential explanations that have been put forth to explain why there is no link found between school spending and student outcomes for cohorts educated since the 1950s include (i) diminished returns to school spending as levels of spending have increased over time (relative to earlier cohorts), (ii) deterioration of the quality of the teaching workforce, and (iii) increased waste and ineffective allocation of resources to school inputs (see Betts 1996).

in student outcomes. This is unlikely to yield causal relationships because many of the changes to how schools have been funded since the 1960s would lead to biases that weaken the observed association between *changes* in school resources and student outcomes. For example, under the Elementary and Secondary Education Act of 1965, school districts that see increasing shares of low-income students over time would receive additional funding. Such policies that link changes in the student population to changes in spending likely generate a negative relationship between school spending and student achievement that would negatively bias the observed relationship between school spending and student outcomes. Additionally, because localities face trade-offs when allocating finite resources, positive effects of endogenous increases in school spending could be offset by reductions in other kinds of potentially productive spending. We overcome the biases inherent in relying on potentially endogenous observational changes in school resources by documenting the relationship between exogenous quasi-experimental shocks to school spending and long-run adult outcomes.

As documented in Murray, Evans, and Schwab (1998), Hoxby (2001), Card and Payne (2002), and Jackson, Johnson, and Persico (2014), the school finance reforms (SFRs) that began in the early 1970s and accelerated in the 1980s caused some of the most dramatic changes in the structure of K–12 education spending in U.S. history. To isolate plausibly exogenous changes in school resources we investigate the effects of changes in per pupil spending, due only to the passage of court-mandated school finance reforms, on long-run educational and economic outcomes. We link data on SFRs and school spending to longitudinal data on a nationally representative sample of children born between 1955 and 1985 and followed into adulthood. These birth cohorts straddle the period in which SFRs were implemented, and thus were differentially exposed to reform-induced changes in school spending depending on place and year of birth.

We use the timing of the court decision mandating reform and the ensuing type of funding reform introduced as exogenous shifters of school spending. Specifically, for each district we predict the spending change that the district would experience after the passage of court-mandated school finance reform based on the experiences of similar districts facing similar reforms in different states. We then see if "exposed" cohorts (those young enough to have been in school during or after the reforms were passed) have better outcomes relative to "unexposed" cohorts (children who were too old to be affected by reforms at the time of passage) in districts predicted (based on the experiences of similar districts in other states) to experience larger reform-induced spending increases. Correlating outcomes with only the predicted reform-induced variation in spending, rather than all actual spending, removes the confounding influence of unobserved factors that may determine actual school spending and affect student outcomes.

In related work, Card and Payne (2002) find that court-mandated SFRs reduce SAT score gaps between low- and high-income students. However, Hoxby (2001) finds mixed evidence on the effect of increased spending due to SFRs on high school dropout rates, and Downes and Figlio (1998) find no significant changes in the distribution of test scores.[2] Looking at individual states, Guryan (2001), Papke (2005), and Roy (2011) find that reforms improved test scores in low-income districts in Massachusetts and Michigan.[3] Overall, the evidence on the effects of SFRs on academic outcomes is mixed, and the effects on long-run economic outcomes is unknown.

Our event study and instrumental variables models reveal that increased per pupil spending induced by SFRs increased the educational attainment and improved the adult labor market outcomes of low-income children. Although we find small effects for children from affluent families, for low-income children, a 10% increase in per pupil spending each year for all 12 years of public school is associated with 0.46 additional years of completed education, 9.6% higher earnings, and a 6.1 percentage point reduction in the annual incidence of adult poverty. The results imply that a 25% increase in per pupil spending throughout one's school years could eliminate the average attainment gaps between children from low-income (average family income of $31,925 in 2000 dollars) and nonpoor families (average family income of $72,029 in 2000 dollars). We present several additional tests that support a causal interpretation. To shed light on mechanisms, we document that reform-induced school spending increases were

2. However, Downes and Figlio (1998) find that plans that impose tax or expenditure limits on local governments reduce overall student performance on standardized tests.

3. Hyman (2014) analyzes the same Michigan reform and finds evidence that it increased college going.

associated with reductions in student-to-teacher ratios, longer school years, and increased teacher salaries—suggesting that improvements in these school inputs improved student outcomes. These findings stand in contrast to studies finding little effect of measured school inputs on student outcomes for cohorts educated after 1950 (Betts 1995, 1996; Hanushek 2001) and are in line with studies that find that school inputs matter for older cohorts educated between 1920 and 1950 (Card and Krueger 1992; Loeb and Bound 1996) and studies on recently educated cohorts using randomized and quasi-random variation in school inputs (e.g., Fredriksson, Öckert, and Oosterbeek 2012; Chetty et al. 2013).

Importantly, we are able reconcile our results with the existing literature by showing (using our data) that observational variation in spending may confound family and neighborhood disadvantage with increased spending, and that districts allocated the additional funds received due to the passage of court-ordered SFRs toward seemingly more productive school inputs than they did endogenous spending increases. We also discuss and highlight the countervailing forces that can explain why there have only been moderate improvements in student outcomes in the past 30 years despite large national increases in per pupil spending.

The remainder of the article is organized as follows. Section II describes the school finance reforms. Section III presents the data used. Section IV outlines our empirical strategy. Section V presents results from both event study and instrumental variables analyses. Section VI presents evidence on the role of specific school resource inputs, and Section VII presents our conclusions. All appendix material is in the Online Appendix.

II. Overview of Court-Ordered Reforms

We aim to document the relationship between long-run outcomes and exogenous variation in school spending. To this aim, we isolate exogenous variation in per pupil school spending caused by the passage of court-ordered SFRs. In most states, prior to the 1970s, most resources spent on K–12 schooling was raised through local property taxes (Hoxby 1996; Howell and Miller 1997). Because the local property tax base is typically higher in areas with higher home values, and there are high levels of residential segregation by socioeconomic status, heavy reliance on local financing contributed to affluent districts' ability to spend more per student. In response to large within-state differences in per pupil spending across wealthy/high-income and poor districts, state supreme courts overturned school finance systems in 28 states between 1971 and 2010, and many states implemented legislative reforms that led to important changes in public education funding. Online Appendix A presents the timing and nature of the court-ordered SFRs in each state.

Challenges to state school finance systems were argued on either equity or adequacy grounds. The early challenges (1971 to mid–1980s) were won on equity grounds. For equity cases, local financing was found to violate the responsibility of the state to provide a quality education to all children. Equity cases sought to weaken the relationship between the quality of educational services and the fiscal capacity of the district. The more recent challenges (late 1980s onward) were mounted on adequacy grounds. Adequacy cases rely on the fact that most states have a constitutional provision requiring the state to provide some adequate level of free education for children (Lindseth 2004) and were argued on the grounds that low per pupil spending levels in certain districts meant that the state had failed to meet this obligation.

Irrespective of the nature of the legal challenges, once the prevailing school finance system was found unconstitutional, most SFRs changed the parameters of spending formulas to reduce inequality in school spending and weaken the relationship between the level of educational spending and the wealth and income level of the district (Card and Payne 2002). The design of state formulas to meet these goals, however, was highly variable. As pointed out in Hoxby (2001), the effect of a SFR on school spending depends on (i) the type of school funding formula introduced by the reform and, (ii) how the funding formula interacts with the specific characteristics of a district. To capture some of this complexity, we follow the typology outlined in Jackson, Johnson, and Persico (2014) and categorize reforms into five main types. Foundation plans guarantee a base level of per pupil school spending and are designed to increase per pupil spending for the lowest-spending districts. Spending limits prohibit per pupil spending levels above some predetermined amount. Such plans tend to reduce spending for high spending and more affluent districts and may reduce spending in the long run for all districts. Reward for effort plans match locally raised funds for education with additional state funds (often with higher match rates for lower-income areas). Such plans will tend to increase spending for all districts with larger increases for districts in lower-income areas. Finally, equalization plans aim to equalize spending levels typically by taxing all districts and redistributing funds to lower-wealth and lower-income districts. Note that these reform types are not mutually exclusive. Online Appendix D details these reform types. These differences in how states implemented SFRs will play a key role in our empirical strategy to isolate exogenous variation in school spending across birth cohorts within a district.

III. Data

We compiled data on school spending, linked them to a database describing various SFRs, and linked these data to a nationally representative longitudinal data set that tracks individuals from childhood into adulthood. Education funding data come from several sources that we combine to form a panel of per pupil spending for U.S. school districts in 1967 and annually from 1970 through 2010.[4] To avoid confounding nominal changes with real changes in spending over time, we convert school spending across all years to 2000 dollars using the Consumer Price Index from the Bureau of Labor Statistics. We use the school district boundaries that prevailed in 1969 to link school districts to counties and pull county-level median family income data from the 1970 census. The spending data are then linked to a database of reforms between 1972 and 2010.[5]

Our data on longer-run outcomes come from the Panel Study of Income Dynamics (PSID) that links individuals to their census blocks during childhood.[6] Our sample consists of PSID sample

4. The Census of Governments has been conducted every five years since 1972 and records school spending for every school district in the United States. The Historical Database on Individual Government Finances (INDFIN) contains school district finance data annually for a subsample of districts from 1967, and 1970 through 1991. After 1991, the CCD School District Finance Survey (F-33) includes data on school spending for every school district in the United States. Additional details on the data and the coverage of districts in these data are in Online Appendix B.

5. A detailed description of how this database of reforms was compiled is in Online Appendix C.

6. The PSID began interviewing a national probability sample of families in 1968. These families were reinterviewed each year through 1997, when interviewing became biennial. All persons in PSID families in 1968 have the PSID “gene,”

members born between 1955 and 1985 who have been followed into adulthood through 2011.[7] These cohorts straddle the first set of court-mandated SFRs (the first court order was in 1971) and are also old enough to have completed formal schooling by 2011. Two-thirds of the sample grew up in a state that was subject to a court-mandated SFR between 1971 and 2000. We match the earliest available childhood residential address to the school district boundaries that prevailed in 1969 to avoid complications arising from endogenously changing district boundaries over time. The algorithm is outlined in Online Appendix E.[8] Each record is merged with data on school spending and the aforementioned school finance variables at the school district level that correspond with the prevailing levels during their school years. Finally, we merge in county characteristics from the 1962 Census of Governments and 1970 census, and information on other key policy changes (described in Section II) during childhood, allowing for an unusually rich set of controls.[9]

The final sample includes 93,022 adult person-year observations of 15,353 individuals (9,035 low-income children; 6,318 nonpoor children) from 1,409 school districts, 1,031 counties, and all 50 states and the District of Columbia. To describe the

which means that they are followed in subsequent waves. When children with the gene become adults and leave their parents' homes, they become their own PSID "family unit" and are interviewed in each wave. The original geographic cluster design of the PSID enables comparisons in adulthood of childhood neighbors who have been followed over the life course. Studies have concluded that the PSID sample remains representative of the national sample of adults (Fitzgerald et al., 1998a,b).

7. We include both the Survey Research Center component and the Survey of Economic Opportunity component, commonly known as the "poverty sample," of the PSID sample.

8. Many school districts were counties during this period, including more than one-half of Southern school districts.

9. The data we use include measures from 1968–1988 Office of Civil Rights (OCR) data; 1960, 1970, 1980, and 1990 census data; 1962–1999 Census of Governments (COG) data; Common Core Data (CCD) compiled by the National Center for Education Statistics; Regional Economic Information System (REIS) data; a comprehensive case inventory of court litigation regarding school desegregation over the 1955–1990 period (American Communities Project); and the American Hospital Association's Annual Survey of Hospitals (1946–1990) and the Centers for Medicare and Medicaid Services data files (dating back to the 1960s) to identify the precise date in which a Medicare-certified hospital was established in each county of the United States (an accurate marker for hospital desegregation compliance).

home environment during childhood, we average parental income and education variables over the ages of 12 and 17 and measure family structure at birth. Following Ben-Shalom, Moffitt, and Scholz (2012) and Short and Smeeding (2012), a child is defined as “low income” if parental family income falls below two times the poverty line for any year during childhood.[10] This captures both the poor and the nearly poor. Henceforth, children from families who were not low income (as defined above) will be referred to as “nonpoor.” The average childhood family incomes for children from low-income and nonpoor families were $31,925 and $72,029 in 2000 dollars, respectively. To compare individuals from different birth cohorts at similar ages, we focus on adult observations between the ages of 20 and 45. The set of adult outcomes examined include (i) educational outcomes—whether graduated from high school, years of completed education (at the most recent survey)—and (ii) labor market and economic status outcomes (measured annually and expressed in 2000 dollars)—wages, family income, and annual incidence of poverty in adulthood (ages 20–45). Summary statistics are presented in Table I.

Average years of completed education is 13.18, and children from low-income families have about 1 year less schooling than the nonpoor. The wage (annual earnings/annual work hours) is our main labor market outcome. We compute the wages only for those who have positive earnings in a given year and are not full-time students. Because we have multiple adult observations for each individual, we have valid wage observations for about 95% of the sample. We show that this feature of the data allows us to better detect effects for those with low labor market attachment. The average wage (in 2000 dollars) at age 30 for those from low-income families is $10.60 and for those from nonpoor families it is $13.60. As one might expect, individuals from more affluent childhood families have higher family incomes and are less likely to be in poverty as adults. We show in Section V that increases in school spending narrow some of these gaps in adult outcomes between those from high- and low-income families.

10. The poverty line is defined by family composition, such that children are defined as “low income” if the family’s income-to-needs ratio falls below 2 for any year during childhood. The income-to-needs ratio is defined using the official federal census poverty thresholds of needs for respective household composition. Due to the oversampling of poor families, 59% of the sample were low-income as children.

TABLE I

DESCRIPTIVE STATISTICS BY CHILDHOOD POVERTY STATUS

	All ($N=15{,}353$)	Low-Income Child ($N=9{,}035$)	Nonpoor Child ($N=6{,}318$)
Adult outcomes (from the PSID):			
High school graduate	0.86	0.79	0.92
Years of education	13.18	12.63	13.64
Ln(wages), at age 30	2.51	2.36	2.61
Adult family income, at age 30	$49,308	$35,212	$55,324
In poverty, at age 30	0.08	0.13	0.04
Age (range: 20–57)	32.9	32.6	33.2
Year born (range: 1955–1985)	1969	1970	1968
Female	0.44	0.43	0.44
Black	0.14	0.23	0.07
Childhood school variables:			
Per pupil spending (avg., ages 5–17)	$4,800	$4,706	$4,873
Any court-ordered SFR, age 5–17	0.53	0.53	0.53
Years of exposure to SFR, age 5–17	4.35	4.46	4.27
1960 district poverty rate (%)	22.09	24.75	19.88
Childhood family variables (from the PSID):			
Income-to-needs ratio (avg., ages 12–17):	3.17	1.64	3.77
Mother's years of education	12.05	11.32	12.66
Father's years of education	12.05	10.91	12.93
Born into two-parent family	0.62	0.55	0.68
Low birth weight (<5.5 pounds)	0.07	0.08	0.06
Childhood neighborhood variables:			
County poverty rate	0.11	0.16	0.08
Residential segregation dissimilarity index	0.72	0.71	0.72

Notes. All descriptive statistics are sample weighted to produce nationally representative estimates of means. Dollars are CPI-U deflated in real 2000 dollars. The income-to-needs ratio is defined using the official federal census poverty thresholds of needs for respective household composition. All adult outcomes and childhood family variable are from the PSID. Per pupil spending data are from the Historical Database on Individual Government Finances (INDFIN) and after 1991, the CCD School District Finance Survey (F-33). Data on the exact timing and type of court-ordered and legislative SFRs was obtained from Public School Finance Programs of the Unites States and Canada (PSFP, see US Department of Education 2001), National Access Network's state by state school finance litigation map (2011), from Murray, Evans, and Schwab (1998), Hoxby (2001), Card and Payne (2002), Hightower, Mitani, and Swanson (2010), and Baicker and Gordon (2004). The district poverty rate is from the 1962 Census of Governments. The childhood poverty rate and dissimilarity index are from the 1960 census.

IV. EMPIRICAL FRAMEWORK

Our goal is to identify the causal effect of per pupil public school spending during childhood on adult outcomes. Because the correlation between per pupil spending in an area and the adult outcomes of students who attended those schools is likely confounded by other factors (due to residential segregation, Tiebout sorting, compensatory spending increases, etc.), we search for exogenous variation in per pupil spending. To this aim, we use only variation in school spending during childhood that can be attributed to the passage of court-ordered SFRs. As discussed in Section II, the goal of SFRs was to increase spending levels in low-spending districts and reduce the differences in per pupil school spending levels across districts. By design, some districts experienced spending increases while others experienced decreases (Murray, Evans, and Schwab 1998; Hoxby 2001; Card and Payne 2002). We use the variation in school spending that resulted from the SFR goal to increase funding in low-spending districts and reduce differences in funding levels across districts. We treat this variation as exogenous and use the resulting natural experiment to estimate the causal effect of per pupil spending on adult outcomes.

To motivate our empirical strategy, we describe the policy experiment below. Individuals who turned 17 years old during the year of the passage of a court-ordered SFR in their state should have completed secondary school by the time reforms were enacted. Such cohorts should be unaffected by the reforms so we classify them as *unexposed*. In contrast, individuals who turned 16 years old or were younger during the year of the passage of a court-ordered SFR would likely have been attending primary or secondary school when reforms were implemented. We refer to these cohorts as *exposed*. One can estimate the exposure effect on adult outcomes for individuals from a particular district by comparing the change in outcomes between exposed and unexposed birth cohorts from that district. To account for any underlying differences across birth cohorts, one can use the difference in outcomes across the same birth cohorts in nonreform districts as a comparison. The difference in outcomes between exposed and unexposed cohorts in a treated district minus the difference in outcomes across the same birth cohorts in comparison districts yields a difference-in-difference (DiD) estimate of the exposure effect on outcomes for that district. Our key identifying assumption is that the spending changes caused by the reforms within districts were unrelated to other district-level changes that could affect adult outcomes directly. Under this assumption, a natural test of whether there is a causal effect of per pupil spending during childhood on adult outcomes is whether the difference in outcomes between exposed and unexposed cohorts from the same school district (i.e., the exposure effect) tends to be larger for those districts that experience larger reform-induced increases in per pupil spending across exposed and unexposed cohorts (i.e., a dose-response effect). An additional test is whether we witness larger improvements in adult outcomes for individuals that experienced those spending increases for more of their school-age years.

We operationalize these intuitive tests using a two-stage least-squares (2SLS) DiD regression model where per pupil spending during childhood is our endogenous treatment variable of interest. As a shorthand, we refer to the change in per pupil spending that occurs within a district because of the passage of a court-ordered SFR as *dosage*. We predict plausibly exogenous spending changes within districts across cohorts using measures of exposure to court-mandated SFRs and measures of exposure to court-mandated SFRs interacted with predictors of dosage.[11] Following the approach of Card and Krueger (1992), our measure of school spending during childhood is the average school spending (in real 2000 dollars) during expected school-age years (ages 5–17) in an individual's childhood school district (hereinafter referred to as spending), $P\overline{P}E_{5-17}$.[12] To quantify the relationship between spending and adult outcomes using only the variation in spending associated with the passage of a court-mandated SFR, we estimate systems of equations of the following form by 2SLS.

$$\ln(P\overline{P}E_{5-17})_{idb} = \pi_1(\text{Exp}_{idb} \times \text{Dosage}_d) + \pi_2(\text{Exp}_{idb}) + \Pi C_{idb} + \rho_d + \rho_b + \xi_{idb} \tag{1}$$

$$Y_{idb} = \delta \cdot \ln(\widehat{P\overline{P}E_{5-17}})_{idb} + \Phi C_{idb} + \theta_d + \theta_b + \varepsilon_{idb}. \tag{2}$$

11. In principle, one could use only the effect of exposure to predict changes in the level of spending. However, using exposure on its own likely violates the monotonicity assumption for a valid instrument because some SFRs lead to increased spending whereas others lead to decreases. Even if exposure alone were a valid instrument, such an approach would exclude all the variation that occurs across districts within a state as a result of the passage of an SFR. Indeed, using only the exposure variation and ignoring variation in dosage yields a weak first stage (F-statistic of 5.7).

12. The average level of district per pupil spending across all school-age years provides a summary measure of the level of financial resources available in the individual's childhood school district during all their school-going years (ages 5–17 corresponding to expected grades K–12). We use the natural log of this average measure to capture the fact that school spending likely exhibits diminishing marginal product (all results are robust to using the level of average school-age spending and are presented in Online Appendix F).

Our endogenous treatment variable $\ln(\overline{PPE}_{5-17})_{idb}$ is the natural log of average school-age spending for individual i from district d in birth cohort b. To only rely on variation across birth cohorts within districts we include district fixed effects ρ_d and θ_d in the first and second stage, respectively. To account for general underlying differences across birth cohorts (irrespective of exposure), we include birth-cohort fixed effects ρ_b and θ_b in the first and second stage, respectively. With the birth-cohort fixed effects, our estimated changes across birth cohorts in reform districts are all relative to the changes across the same birth cohorts in districts that did not implement reforms during that time. Our measure of exposure, Exp_{idb}, is the number of school-age years occurring after the passage of a state court-ordered SFR for individual i in birth cohort b from district d. Exp_{idb} varies at the state–birth cohort level. This variable goes from 0 (for those who turned age 17 or older the year of the state's court order) to 12 (for those who turned age 5 or were younger the year of the state's court order). To capture variation in dosage conditional on exposure, we interact Exp_{idb} with $Dosage_d$. $Dosage_d$ is a district-level measure of the amount of spending change caused by the court-ordered SFR in district d. We detail exactly how $Dosage_d$ is measured in Section IV.A. ξ_{idb} and ε_{idb} are random error terms.

This DiD 2SLS model compares the difference in outcomes between birth cohorts from the same district exposed to reforms for different amounts of time (variation in exposure) across districts with larger or smaller reform-induced changes in per pupil school spending (variation in dosage). If exposed cohorts from districts that experience larger reform-induced spending increases also tend to subsequently experience larger improvements in adult outcomes (relative to unexposed cohorts) then the coefficient δ from equation (2) will be positive (for an outcome such as wages for which larger positive values are better). However, the coefficient δ will be 0 if exposure to larger reform-induced spending changes (across cohorts) are unrelated to changes in adult outcomes. As long as the timing of court-mandated SFRs is exogenous to changes in outcomes across birth cohorts within districts, the coefficient δ should uncover the causal effect of school spending on adult outcomes. It is important to note that because the childhood school district prior to reforms may not always be the same school district an individual actually attends (due to residential mobility after reforms), δ is an intention-to-treat estimate that quantifies the policy effect of increasing per pupil school spending in an individual's childhood school district.[13]

A key variable in our analysis is $Dosage_d$, the spending change experienced by exposed cohorts from district d because of the court-mandated SFR. Even though $Dosage_d$ is not directly observed, one can use proxies (or predictors) of dosage in equation (1). Although using the actual change in spending experienced by a district after reforms would seem like a reasonable proxy for dosage, we do not take this approach to avoid endogeneity bias. The actual change in spending experienced by a district after reforms would include all changes in spending that happen to coincide with the timing of the court order. To avoid using changes in school spending that may be the result of other policy changes within the district or other local changes that could directly affect outcomes, we predict dosage in district d using only characteristics of the district prior to the initial court-ordered SFR. This excludes any changes in school spending that might have been caused by other district-level changes that could directly affect student outcomes. We propose two prereform predictors of dosage. We detail each predictor in Section IV.A. We also show that DiD 2SLS estimated spending effects are similar using either predictor.

To ensure that we isolate changes due to court-mandated SFRs, we include C_{idb}, a vector that includes a variety of additional individual, family, and childhood county controls. These include parental education and occupational status, parental income, mother's marital status at child's birth, birth weight, child health insurance coverage, and gender. C_{idb} also includes race-by-census-division birth-cohort fixed effects, and birth-cohort linear trends interacted with various 1960 characteristics of the childhood county (poverty rate, percent black, average education, percent urban, and population size). Finally, to avoid confounding our effects with that of other policies that overlap our study period, C_{idb} includes controls for county-by-year

13. Because some individuals may have moved away from their prereform school district or may have dropped out of school before the age of 17, our measure of school-age spending is a noisy measure of the school spending individuals were actually exposed to. Using the actual spending an individual is exposed to would introduce selection bias, because the level of spending would be determined in part by the decisions of individual parents. By using the individual's childhood residential location prior to the court order, one removes any bias due to endogenous residential sorting.

measures of school desegregation, hospital desegregation, community health centers, state funding for kindergarten, per capita Head Start spending, Title I school funding, imposition of tax limit policies, average childhood spending on food stamps, Aid to Families with Dependent Children, Medicaid, and unemployment insurance (Chay, Guryan, and Mazumder 2009; Hoynes, Schanzenbach, and Almond 2012; Johnson 2011). Standard errors are clustered at the school district level.[14]

Underlying this 2SLS DiD model is a first-stage DiD model that predicts changes in per pupil spending for exposed cohorts that are more positive in districts with higher dosage. The credibility of our research design hinges on the assumption that the timing of court-ordered SFRs were unrelated to other district-level changes that directly influence outcomes (irrespective of dosage). Accordingly, in the interest of transparency, in the next section we describe our two proposed predictors of dosage, and we present flexible DiD event study effects of the initial court-ordered SFR on spending by different levels of our dosage measures. Although there is no perfect test of the assumption that the reform-induced spending changes were exogenous, the event study figures based on different levels of predicted dosage lay bare the policy variation underlying our 2SLS DiD approach and allow one to visually assess the credibility of our research design.

IV.A. Creating Measures of Dosage Based on Prereform Characteristics

As described already, a key step in our empirical strategy is to identify those school districts that should experience larger versus smaller spending changes due to reforms (i.e., identify districts that experience differences in dosage conditional on the exogenous timing of exposure to court-ordered SFRs). We propose two such approaches that we discuss and analyze below.

Approach 1. Prior studies have found that court-ordered SFRs tend to equalize per pupil school spending within states by increasing spending for previously low-spending districts with

14. All results are robust to clustering the standard errors at the childhood state level. Our main results that cluster the standard errors by state are presented in Online Appendix G.

small effects for previously high-spending districts (Murray, Evans, and Schwab 1998; Card and Payne 2002). As such, our first approach to identifying districts that on average would experience larger versus smaller spending increases after a court-ordered SFR (without using any potentially endogenous changes that actually occurred in that district around the time of reforms) is to use the relative spending level of the district prior to the court-ordered SFR.

To show visually that the effect of court-ordered SFRs on the changes in level of school spending experienced by an individual during their school-age years varies by the prereform per pupil spending levels of their childhood district, we estimate an event study model based on the DiD first-stage model described in equation (1). Specifically, we estimate a flexible version of the first-stage equation (1) where our predictor of dosage is the quartile of the district in the state distribution of per pupil spending in 1972. Note that the first court order was issued in 1971 and enacted in 1972 so that the 1972 fiscal year (1971–1972 academic year) is the last prereform year. To map out the change in per pupil spending for cohorts that attended primary and secondary school before, during, and after the passage of a court-ordered SFR, we replace the linear measure of exposure, $Exp_{idb,}$ with a series of indicator variables denoting the number of years after the individual turned 17 that the court order occurred. Specifically, we implement this DiD event study by estimating equation (3) by ordinary least squares (OLS).

$$\ln\,(P\overline{P}E_{5-17})_{idb} = \sum_{Q_{ppe}=1}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}}\right)\cdot \alpha_{T,Q_{ppe}} + \Pi C_{idb} + \theta_d + \theta_{brg} + \upsilon_{idb} \quad (3)$$

All common variables are as in equation (1), and $Q_{ppe72,d}$ are indicators for the quartile of district d in the state distribution of per pupil spending in 1972. These are time-invariant district characteristics that describe whether district d was high- or low-spending prior to reforms, and function as our key exogenous predictors of dosage. Because some states had multiple court-mandated SFRs, for simplicity, we estimate treatment effects only for the first court-mandated SFR. The variable T_{idb} is the year individual i from school district d turned age 17 minus the year of the initial SFR court order in school district d.[15] Accordingly, the timing indicators, $I_{T_{idb}=T}$, equal 1 if the year individual i from school district d turned age 17 minus the year of the initial SFR court order in school district d equals T and 0 otherwise. We include indicators for values of T between -20 and 20. Values of T between -20 and -1 represent unexposed cohorts who turned between the ages of 18 and 37 in the year of the initial court order; a value of 0 is our reference category and represents individuals who turned 17 in the year of the initial court order and were thus not exposed; values between 1 and 11 represent exposed cohorts who were "partially treated" because they were of school-going age (6–16) at the time of the initial court order but had less than 12 years of expected exposure; and values of 12 and greater represent fully treated exposed cohorts who turned 5 or were younger during the year reforms were enacted and were therefore expected to attend all 12 years of public schooling during postreform years.

Each of the event time indicator variables is interacted with four indicators denoting the quantile of the childhood district in the state distribution of per pupil spending in 1972, $I_{Q_{ppe72,d}=Q_{ppe}}$. Accordingly, the coefficients for the two-way interactions, $\alpha_{T,Q_{ppe}}$, map out the dynamic treatment effects (across birth cohorts from the same school district) of the first court-ordered SFR on log average school-age school spending for individuals from districts in spending quartile Q_{ppe}.[16] We plot the estimated dynamic treatment effects to illustrate how spending evolves for cohorts in school before, during, and after reforms (relative to changes for the same birth cohorts in similar districts in nonreform states). These estimates illustrate the exact timing of changes in school-age spending in relation to the number of school-age years of exposure to the court-ordered SFR for individuals from school districts with high versus low prereform per pupil spending. A plot of the coefficients $\alpha_{T,Q_{ppe}}$ across prereform spending quartiles is a visual depiction of our first stage isolating reform-induced

15. The indicator for event time 20 includes all years with event time above 20. Similarly, the indicator for event time -20 includes all years with event time less than minus 20.

16. For example, $\alpha_{-10,1}$ is the effect of the passage of a court-ordered SFR on the school-age per pupil spending of the untreated cohorts that turned age 17 10 years prior to reforms from districts in the first (bottom) prereform spending quartile. Also, $\alpha_{-5,2}$ is the effect of the passage of a court-ordered SFR on the school-age per pupil spending of the treated cohorts that turned age 17 five years after reforms from districts in the second prereform spending quartile.

variation in school-age per pupil spending based on variation in exposure and its interaction with prereform spending levels.

Figure I presents the estimated event study plots, $\alpha_{T,Q_{ppe}}$, for different quartiles of prereform spending. We find that all districts other than the top quartile of spending experienced sizable increases in per-pupil spending. Thus, for the sake of parsimony, we present the event study graphs for districts in the top quartile (Top) and that for the bottom three quartiles (Bottom). The figure depicts how school-age per pupil spending evolved for cohorts that were expected to graduate 7 years prior to the first court-mandated reform through those that were expected to graduate 17 years postreform. Each series of event study estimates is relative to the effect for year 0 (those that turned 17 in the year of the first court-ordered SFR in their state). Because the outcome is in logs, the values represent percent changes in average school-age spending relative to the cohort from the same district that was 17 the year of the first court-ordered SFR. We present the effect of court-ordered SFRs on spending for the sample of districts linked to individuals in the PSID. Similar plots using all districts are presented in Online Appendix B.

As one can see, unexposed cohorts -7 through -1 (turned ages 18–24 the year of the first court order) in both high- and low-spending districts in reform states saw similar changes in school-age per pupil spending as districts with the same prereform spending level in nonreform states (or other nonreform years in reform states). The p-value for joint hypothesis that all these prereform event study years is equal to 0 for both the high- and low-spending group is above .1. The fact that districts in all quartiles of 1972 spending in reform states were on a similar trajectory as districts in nonreform states shows that districts that are expected to experience increases in school spending due to reforms were not already on a differential trajectory of improving outcomes. This lends credibility to the exogeneity of reform-induced spending changes.

Consistent with court-ordered SFRs reducing spending inequality, exposed cohorts in initially lower-spending districts (bottom panel) see large spending increases that increase with years of exposure, while the highest-spending districts experience small increases. Among those with 12 years of exposure (age five during the year of the initial court order), those from high-spending districts experienced a 6 percent increase in average school-age spending while those in low-spending districts experienced a 12

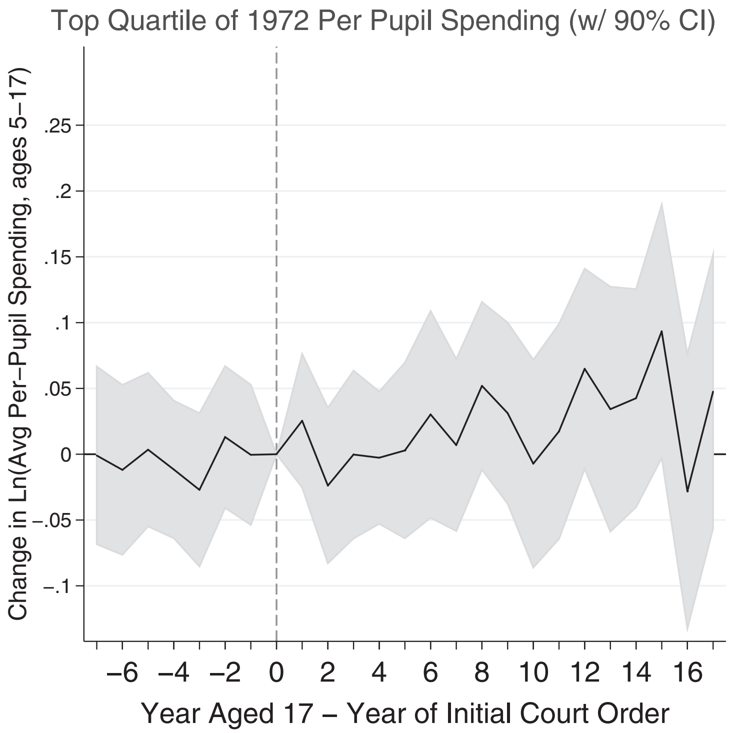


FIGURE I

The Effect of a Court-Ordered Reform on School-Age Per Pupil Spending by Prereform Spending Quartile

Data: PSID geocode data (1968–2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011 ($N = 15{,}353$ individuals from 1,409 school districts [1,031 child counties, 50 states]). Sampling weights are used so that the results are nationally representative.

Continued.

percent increase. A test of equality of the postyear indicators across the two groups yields a p-value below .05 such that the spending level in a district prior to reforms relative to others in the same state is a valid exogenous predictor of dosage.

Approach 2. Hoxby (2001) demonstrates that the effect of an SFR on spending depends on the type of reforms implemented so that among high- or low-spending districts, there are differences

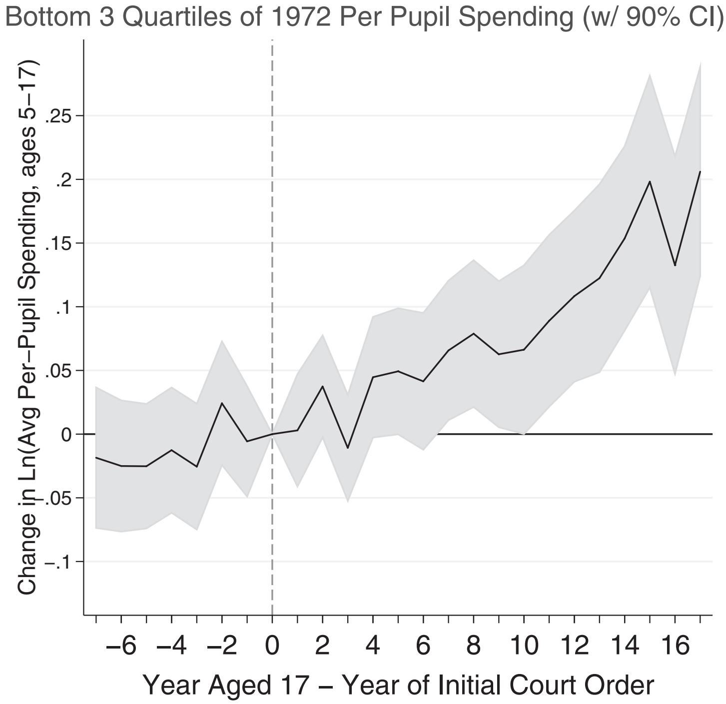


FIGURE I

Models: The event study plot is based on indicator variables for the number of school-age years of exposure to a court-ordered SFR interacted with an indicator for whether the district was in the top quartile of the state distribution of per pupil spending in 1972. Results are based on nonparametric event study models that include school district fixed effects, race × census division × birth cohort fixed effects, and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race) each interacted with linear cohort trends.

in dosage due to differences in reform type. Also, Card and Payne (2002) demonstrate that many reforms reduced the strength of the relationship between district income levels and per pupil spending rather than simply reducing the dispersion of spending per se. As such, among high- or low-spending districts, there are differences in dosage due to prereform differences in district income levels.[17] It follows that knowing both the prereform spending and income level of the district in addition to the type of reforms introduced by the court order may allow one to better predict the reform-induced spending change for individual districts. Thus, we propose a finer-grained predictor of dosage that incorporates additional information about (i) the income level of the district prior to the court order and (ii) the type of funding formulas introduced in response to the court order. To this aim, we create a scalar district-level predictor of dosage that incorporates this additional information.

In the simplest of terms, our second district-specific predictor of dosage, $Spend_d$, is the estimated dosage for observationally similar districts located in other states. By observationally similar we mean districts (in other states) with the same relative spending level prior to reforms, same relative income level prior to reforms, and facing the same funding formula changes as a result of the court order. Using a "leave-out" estimate based on estimated spending changes excluding all data from the own state avoids the mechanical endogeneity of using actual changes to obtain predicted changes. More important, the leave-out estimate excludes any endogenous spending changes that may have occurred in district d around the time of the initial court order when forming our predicted dosage measure. This leave-out predicted dosage serves as a jackknife instrumental variable (Angrist, Imbens, and Krueger 1999) and is constructed as follows:

(i) We exclude all data for the state that includes district d.
(ii) We associate each court-ordered SFR to the reform types caused by the court order. We use the five key reform types described above: foundation plans, spending limits, reward for effort plans, equalization plans, and equity cases. For each court order, we determine the funding formula type the reform introduced by associating any formula changes (there may have been more than one) within three years of a court order to that court order.[18]

17. This is because not all low-income districts are also low-spending, not all high-spending districts are high-income, and there are important differences across states in reform type.

18. Because some formula changes occur as a result of legislative actions unrelated to court rulings, not all formula changes were associated with a court mandate.

(iii) We use the median family income in 1969 (prior to any court-ordered SFRs) for the county associated with each district as our measure of district income. Using this measure, we compute the quartile of each district's median income in 1969 (within the relevant state distribution).

(iv) We augment equation (3) to include years of exposure indicators interacted with the quartile of district median income (as described already) prior to reforms, each interacted with indicators for whether each of the five reform types was introduced as a result of the court order.[19] Using district-by-birth-cohort data for the full universe of districts (but excluding districts in the same state as district d), we estimate equation (6), where all variables are defined as in equation (1), $I_{F,d}$ is an indicator for the type of reform (F) introduced by the court order in the state containing district d, and $Q_{inc69,d}$ is the quartile of district d in the state distribution of median income in 1969.

$$\ln(P\overline{P}E_{5-17})_{idb} = \sum_{Q_{ppe}=1}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\right)\cdot\alpha_{T,Q_{ppe}}$$

$$+\sum_{F=1}^{5}\sum_{Q_{inc}=1}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{inc69,d}=Q_{inc}}\times I_{F,d}\right)\cdot\alpha_{T,Q_{inc},F}+\Pi C_{idb}++\theta_d+\theta_{brg}+\upsilon_{idb}$$

(4)

As in equation (3), the coefficients $\alpha_{T,Q_{ppe}}$ map out the effect of T years of exposure to a court-ordered SFR for those from districts in the Q^{th} quartile of the state distribution of per pupil spending in 1972. Similarly, the coefficients $\alpha_{T,Q_{inc},F}$ map out the effects on school-age per-pupil spending of T years of exposure to a court-ordered SFR that introduced reform type F for those from districts in the Q^{th} quartile of the state distribution of median income in 1969.

(v) Using the estimates from equation (4), for each district in the excluded state from step i, we compute the average

19. Note that models that predict spending changes using any number of interactions between reform type, quartile of district spending in 1972, and quartile of income in 1969 yield very similar regression results. However, the preferred model yields the strongest first-stage and the most precise 2SLS regression estimates.

change in spending for exposed cohorts who were between the ages of 10 and 15 during the year of the initial court order (relative to unexposed cohorts) based only on (a) the reform types/funding formulas introduced by the court order, (b) the quartile of the district in the state distribution of spending prior to the court order, and (c) the quartile of the district in the state distribution of median family income prior to the initial court order.[20]

(vi) Repeat steps i through iv for each state.[21]

In words, $Spend_d$, our district-specific predictor of dosage, is the estimated reform-induced change in school-age spending experienced by those who were between the ages of 10 and 15 in the year of the first court-mandated reform, where the predicted change is based on the experiences of districts in other states with the same prereform relative income level, the same prereform relative spending level, and facing the same kinds of reforms as district d. Because different kinds of reforms may affect districts differently (Hoxby 2001) and many funding formulas are based on a district's spending and income levels (Card and Payne 1999), using additional information about the type of reform and district income level to predict dosage may lead to additional

20. We use the predicted spending change for those who were between the ages of 10 and 15 in the year of the initial court-ordered SFR. As such, in notation form, our predicted effect from equation (2) using data from all *other* states is

$$Spend_d = \frac{\left(\sum_{Q_{ppe}=1}^{4} \sum_{T=2}^{7} \left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}} \right) \cdot \hat{\alpha}_{T,Q_{ppe}} + \sum_{F=1}^{5} \sum_{Q_{inc}=1}^{4} \sum_{T=2}^{7} \left(I_{T_{idb}=T} \times I_{Q_{inc69,d}=Q_{inc}} \times I_{F,d} \right) \cdot \hat{\alpha}_{T,Q_{inc},F} \right)}{6}.$$

Our chosen age range to form this prediction is informed by the fact that in Figure I there is no reform effect on spending for those exposed for only one year and the effect of reforms on spending becomes apparent within seven years of exposure. In principle, we could have chosen any age range between 5 and 17. However, our predictors of dosage are essentially invariant to the age range chosen; the correlation between predicted dosage using ages 10–15 and using ages 5–17 is 0.98. As direct evidence that our conclusions are not sensitive to the specific age range chosen, point estimates from 2SLS models that do not use the leave-out approach (Approach 1) are very similar to those that use the leave-out approach (Approach 2).

21. Note that equation (4) involves estimation of several hundred coefficients. By summarizing the results of this large equation with the fitted value one avoids a many weak instruments problem (Angrist, Imbens, and Krueger 1999). Also, Angrist, Imbens, and Krueger (1999) show that the standard errors computed in Stata for the 2SLS estimator using a leave-out jackknife instrumental variable are very close to those that account for estimation error explicitly.

variation in predicted dosage among districts with the same prereform spending level.

To show the similarities and differences between $Q_{ppe72,d}$ and $Spend_d$, Online Appendix H shows the cross-tabs for the quartiles of district $Spend_d$, spending in 1972, and median income in 1969. As one would expect, areas in the top quartile of $Spend_d$ are disproportionately lower income and lower spending prior to reforms. Similarly, areas in the bottom quartile of $Spend_d$ tend to be higher income and higher spending prior to reforms. However, $Spend_d$ measures variability in predicted dosage that is not captured by prereform spending; only 50% of districts in the bottom quartile of $Spend_d$ are in the top quartile of prereform spending, and only 39% of districts in the top quartile of $Spend_d$ are in the bottom quartile of prereform spending. If this additional variability associated with income levels and reform type picks up real variation in dosage, then $Spend_d$ should be a finer-grained predictor of dosage than prereform spending levels alone.

To illustrate the potential benefits of this additional variability in predicted dosage, in Figure II we plot event study estimates analogous to equation (3) where we replace the four $Q_{ppe72,d}$ indicator variables with a single dichotomous variable that is equal to 1 if $Spend_d$ is positive and 0 otherwise. This indicator denotes whether, based on the experiences of districts in other states with similar prereform characteristics that face the same kinds of reforms, a district is expected to experience a spending increase due to reforms. Roughly two-thirds of districts in reform states are predicted to experience spending increases due to court-ordered SFRs.[22] We separately plot the flexible event study estimates for districts with predicted reform-induced spending increases ($Spend_d > 0$) and those with no predicted spending changes or predicted spending decreases ($Spend_d \leq 0$). The reference cohort is those who turned 17 in the year of the initial court order. If $Spend_d$ identifies clean variation in dosage, then (i) there will be no differential pretrends for unexposed cohorts from either group of districts, and (ii) among exposed cohorts, spending increases

22. Districts predicted to increase spending were predicted to increase by 10% due to the reforms, on average. Districts predicted to decrease spending were predicted to decrease by 8% due to the reforms, on average. As shown in Figure II, the relationship between predicted increases and actual increases is monotonic but nonlinear. This motivates our flexible parameterization of predicted spending increases.

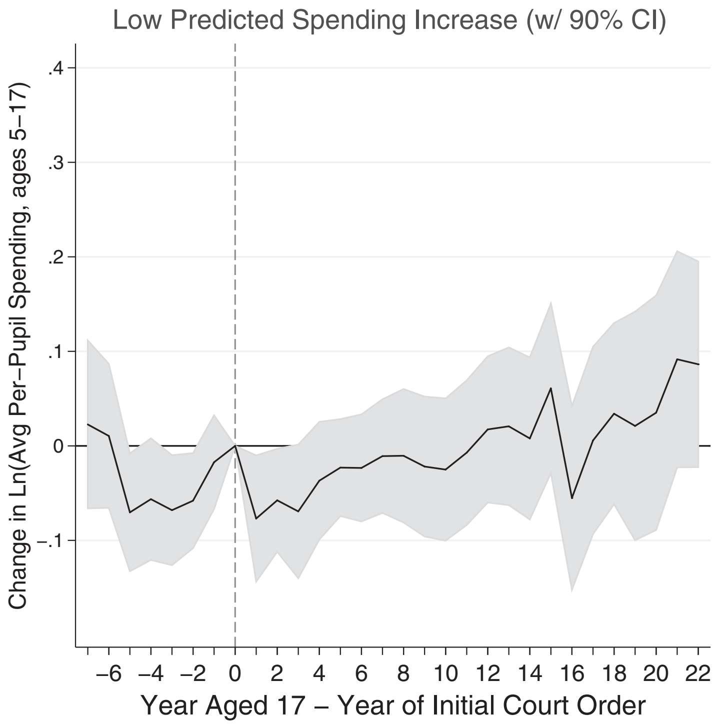


FIGURE II

The Effect of a Court-Ordered Reform on School-Age Per Pupil Spending by Predicted Dosage

High predicted spending increase refers to districts in reform states with $Spend_d > 0$ and low predicted spending increase refers to districts in reform states with $Spend_d \leq 0$. Roughly two thirds of districts in reform states had $Spend_d > 0$.

Continued.

for districts with $Spend_d > 0$ would be greater than those for other districts. We document precisely these patterns.

In Figure II, consistent with the timing of court-ordered SFRs being exogenous to underlying trends in school spending, districts with lower- and higher-predicted dosage were on similar prereform trajectories as similar districts in nonreform states. Consistent with $Spend_d$ isolating real variation in dosage, cohorts that turned five years old during the year of the initial court order (cohort 12) in districts with $Spend_d > 0$ experience a 12 percent increase in school-age per pupil spending

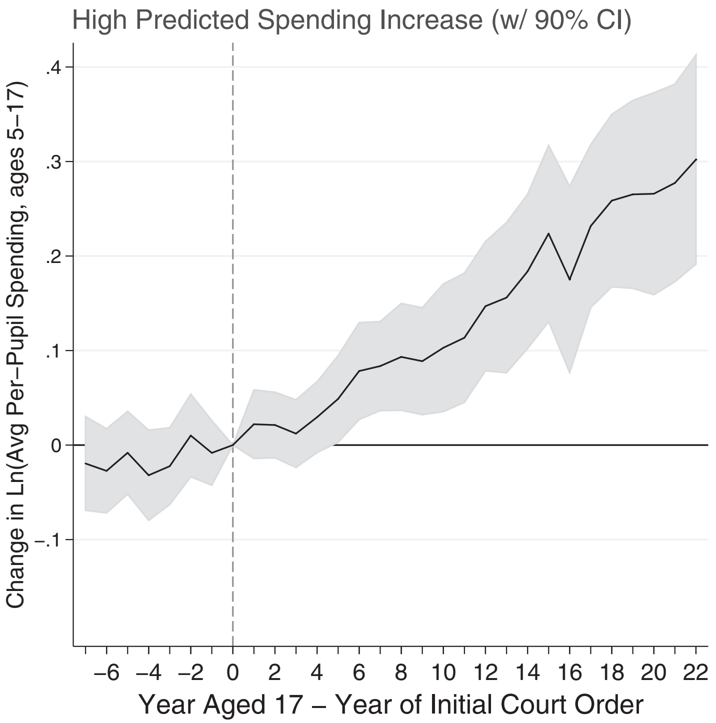


FIGURE II

Data: PSID geocode data (1968–2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011 ($N = 15{,}353$ individuals from 1,409 school districts [1,031 child counties, 50 states]). Sampling weights are used so that the results are nationally representative.

Models: The event study plot is based on indicator variables for the number of school-age years of exposure to a court-ordered SFR interacted with whether the district is predicted to experience a spending increase due to reforms ($Spend_d. > 0$) or not. Results are based on nonparametric event study models that include school district fixed effects, race × census division × birth cohort fixed effects, and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race) each interacted with linear cohort trends.

while the same cohorts in districts with $Spend_d \leq 0$ experience a 4% decrease. The difference in school-age spending for these cohorts across the high- and low-predicted dosage districts is 16% (p-value $< .01$)—more than twice as large as the difference in school-age spending between the top-spending quartiles and other districts for these same cohorts. This increased ability to detect differences in reform-induced spending changes across districts improves our ability to detect outcome differences across these districts.

To assuage any concerns regarding the more complicated leave-out approach (such as the type of funding formula change not being exogenous) we present separate 2SLS regression results using quantiles of $Spend_d$ as a predictor of dosage and results using quantiles of per pupil school spending in 1972 as a predictor of dosage.[23] Both measures of dosage yield very similar results. As such, in the interest of brevity, we focus our discussion on the more refined measure.

IV.B. Potential Biases from Using Observational Variation in School Spending

Our emphasis on using only exogenous variation in spending is motivated by the observation that simply comparing outcomes of students exposed to more or less school spending, even within the same district, could lead to biased estimates of the effects of school spending if there were other factors that affect both outcomes and school spending simultaneously. For example, a decline in the local economy could depress school spending (through home prices or tax rates) and have deleterious effects on student outcomes through mechanisms unrelated to school spending, such as parental income. This would result in a spurious positive correlation between per pupil spending and child outcomes. Conversely, an inflow of low-income, special needs, or English-language learner students could lead to an inflow of compensatory state or federal funding while simultaneously generating reduced student outcomes. This would lead to a spurious negative relationship between spending and student outcomes.

To highlight this point, we test the exogeneity of school spending. First, we predict both high school graduation and adult wages (at age 30) using the fitted values of a regression of

23. Online Appendix I presents event study graphs for outcomes using 1972 spending quartile as the measure of dosage.

these outcomes on parental income, race, mother's and father's education and occupational prestige index, mother's marital status at child's birth, birth weight, childhood county-level average per capita expenditures on Head Start, AFDC, Medicaid, and food stamps during school-age years—this is an effect size weighted index of childhood family/community factors. In Table II, we examine whether predicted outcomes are related to the average district per pupil spending during ages 5–17. Naive OLS models that rely on variation in school spending both within and across states (top panel, columns (1) and (2)) show a strong positive and statistically significant association between school spending and predicted outcomes. This is consistent with most people's priors that raw correlations between spending and outcomes are likely to be positively biased because areas with higher levels of school spending (in the cross-section) will tend to comprise children from more advantaged family backgrounds. However, when we examine the relationship between changes in actual spending within districts over time and changes in predicted outcomes (columns (3) and (4)), there is a statistically significant negative relationship for predicted high school completion and a marginally statistically significant negative relationship for predicted wage at age 30. This is consistent with there being a negative bias when using actual spending changes within districts over time to predict better outcomes. We also look at the relationship between school inputs (student-teacher ratios) and endogenous changes in school spending (column (5)). Surprisingly, although the point estimates show the expected sign, endogenous spending changes are not significantly related to observable school resource inputs.

In contrast to OLS estimates, 2SLS estimates that use only reform-induced school spending changes are not related to changes in predicted outcomes (based on an effect size weighted index of childhood family/community factors), and the point estimates go in different directions for the two predicted outcomes (lower panel). Looking to the student-teacher ratio, however, reveals a stark difference between the identifying OLS variation and the 2SLS variation; reform-induced spending increases are associated with large, statistically significant reductions in the student-teacher ratio. Table II illustrates that OLS estimates of the effects of school spending on outcomes may be negatively biased and may not be associated with improved school inputs. In contrast, the exogenous variation in spending due only to

TABLE II

EFFECT OF ENDOGENOUS AND EXOGENOUS SPENDING INCREASES ON PREDICTED OUTCOMES AND SCHOOL RESOURCES

	(1)	(2)	(3)	(4)	(5)
	Predicted Prob (High School Grad)	Predicted Ln (Wage) at age 30	Predicted Prob (High School Grad)	Predicted Ln (Wage) at age 30	Pupil-Teacher Ratio, age 5–17
	Birth Cohorts and Gender Controls		Birth Cohorts and Gender Controls with District Fixed Effects		Birth Cohorts and Gender Controls with District Fixed Effects
Model: OLS					
$PPE_{d(\text{age 5-17})}$	0.0164***	0.0358***	−0.0044**	−0.0069+	−0.0331
	(0.0022)	(0.0043)	(0.0019)	(0.0048)	(0.0482)
$\text{Ln}(PPE_d)_{(\text{age 5-17})}$	0.0959***	0.1844***	−0.0152+	−0.0223	−0.4052
	(0.0083)	(0.0174)	(0.0111)	(0.0205)	(0.2929)
Model: 2SLS/IV					
$PPE_{d(\text{age 5-17})}$			−0.0022	0.0108	−0.4484**
			(0.0084)	(0.0139)	(0.2569)
$\ln(PPE_d)_{(\text{age 5-17})}$			−0.0074	0.0594	−2.6883**
			(0.0448)	(0.0746)	(1.3838)
Number of individuals	15,353	13,183	15,353	13,183	12,532
Number of school districts	1,409	1,395	1,409	1,395	1,277

Notes. The key treatment variable, $\ln(PPE_d)_{(\text{age 5-17})}$, is the natural log of average school-age per pupil spending. Robust standard errors are in parentheses (clustered at school district level). *** $p < .01$, ** $p < .05$, * $p < .10$; Data: PSID geocode data (1968–2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative. Outcomes: predicted wages and predicted high school graduation are based on models that predict these adult outcomes using only childhood family/community SES characteristics (including parental income, race, mother's and father's education and occupational prestige index, mother's marital status at child's birth, birth weight, childhood county-level average per capita expenditures on Head Start, AFDC, Medicaid, food stamps during school-age years). These predicted outcomes are a weighted index of childhood family/community SES factors. 2SLS: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$).

SFRs is likely to uncover the true causal relationship as mediated by improved school inputs. We show evidence of this in Sections V and VI.

V. Effects on Longer-Run Outcomes

Figures I and II show two distinct sources of variation in spending during one's school-age years: (i) variation in the duration of exposure to spending increases across cohorts from the same district driven by differences in the year of birth relative to the year of the initial court order, and (ii) variation in the size of the spending increase experienced within exposed cohorts across districts driven by the fact that some districts experienced larger reform-induced spending increases than others. Accordingly, there are two natural tests of whether reform-induced spending changes have a causal effect on adult outcomes. The first test is whether exposed cohorts from districts that experienced increases in spending also had improved outcomes relative to unexposed cohorts from the same district. The second test is whether the improvements observed for exposed cohorts (relative to unexposed cohorts) are larger for those from districts that experienced larger spending increases. We implement these tests within an event study framework and present the results graphically.

Toward this goal, before discussing the regression results we present event study estimates similar to Figure II where the dependent variables are the long-run adult outcomes. We present the estimated event study plots on educational attainment and labor market outcomes for individuals from treated districts with predicted reform-induced spending increases ($Spend_d > 0$) and other treated districts ($Spend_d \leq 0$). This graphically presents the reduced-form effect of court-mandated SFRs on outcomes by both duration of exposure and predicted dosage. If there is a causal effect of spending on outcomes, and the spending increases due to reforms are exogenous to changes in outcomes, then (i) the trajectory of outcomes among unexposed cohorts should be similar for those individuals from districts that experience large and small spending increases; (ii) among exposed cohorts from districts that experience spending increases, outcomes should be improving in years of exposure to reforms; and (iii) the effect of exposure should be greater in districts with larger predicted increases in spending (dosage). We present visual evidence of such patterns.

V.A. *Educational Attainment*

Figure III presents the event study estimates of the effects of reform-induced changes in per pupil spending on years of completed education. On the top we present the estimated event study plots for individuals from treated districts with low predicted spending increases ($Spend_d \leq 0$), and on the bottom we present the estimated event study plots for those from treated districts with high predicted spending increases ($Spend_d > 0$). The reference cohort (event study year 0) are those who were 17 at the time of the court decision mandating reform. Each panel shows the within-district dynamic treatment effect of a court-mandated SFR across birth cohorts by predicted dosage level ($Spend_d$) along with the 90% confidence interval for each event study year.

Overall, there is a clear pattern of improved outcomes for exposed cohorts from districts with larger predicted dosage. Among unexposed cohorts (i.e., those that were 17 or older at the time of the reforms), there is no discernible differential trending in educational attainment by predicted dosage. Importantly, the event study estimates for unexposed individuals from both groups of districts hover around 0 (the implied effect for those from nonreform districts), indicating that the timing of the reforms was likely exogenous to changes in educational attainment in a given district and that the size of the predicted spending increase was unrelated to prereform trends in outcomes. This lends credibility to our research design and the resulting 2SLS estimates. Looking at exposed cohorts, the results are consistent with significant causal effects on exposed cohorts that experienced increases in school-age per pupil spending.

In districts with larger predicted spending increases, cohorts with more years of exposure have higher completed years of education than unexposed cohorts and cohorts with fewer years of exposure. Even though each event study year is estimated with noise, among cohorts with more than 5 years of exposure (i.e., those age 12 or younger at the time of the initial court order) the 90 percent confidence interval for most individual event study years lies above 0. Note that testing the difference between individual years of exposure is low powered and is not a test of the

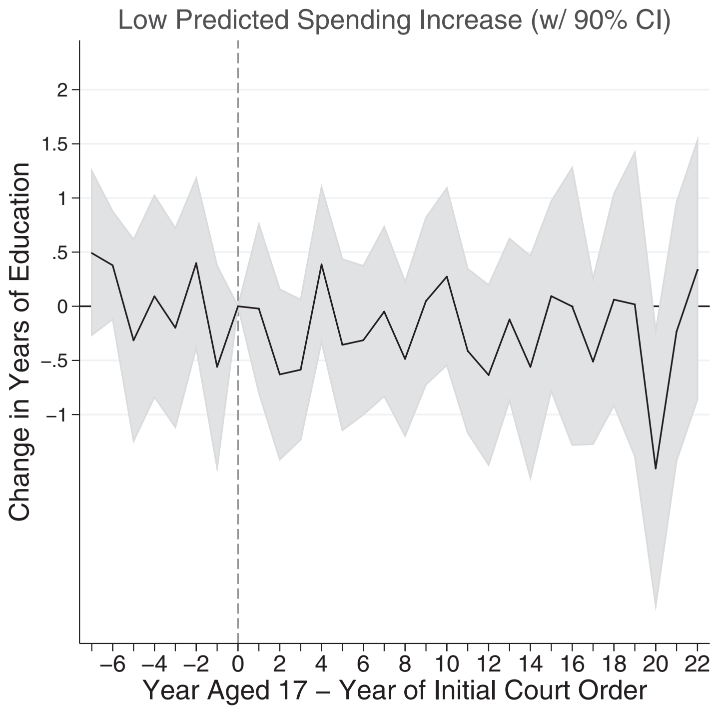


FIGURE III

The Effect of a Court-Ordered Reform on Years of Educational Attainment by Predicted Dosage

High predicted spending increase refers to districts in reform states with $Spend_{d}. > 0$ and low predicted spending increase refers to districts in reform states with $Spend_{d}. \leq 0$. Roughly two thirds of districts in reform states had $Spend_{d}. > 0$.

Continued.

broader hypothesis that variation in school spending is related to variation in outcomes. To test this broader hypothesis, we rely on the 2SLS regressions. Also consistent with a causal impact of school spending, among treated districts with low predicted spending increases that saw either no effect or small decreases in school spending, there is no discernible pattern across exposed cohorts (indicating little effect on educational attainment among exposed cohorts where there was little change in spending). This is further evidenced by that fact that among treated districts with low predicted spending increases, only 1 of the 22 postreform event study year estimates is statistically significantly different

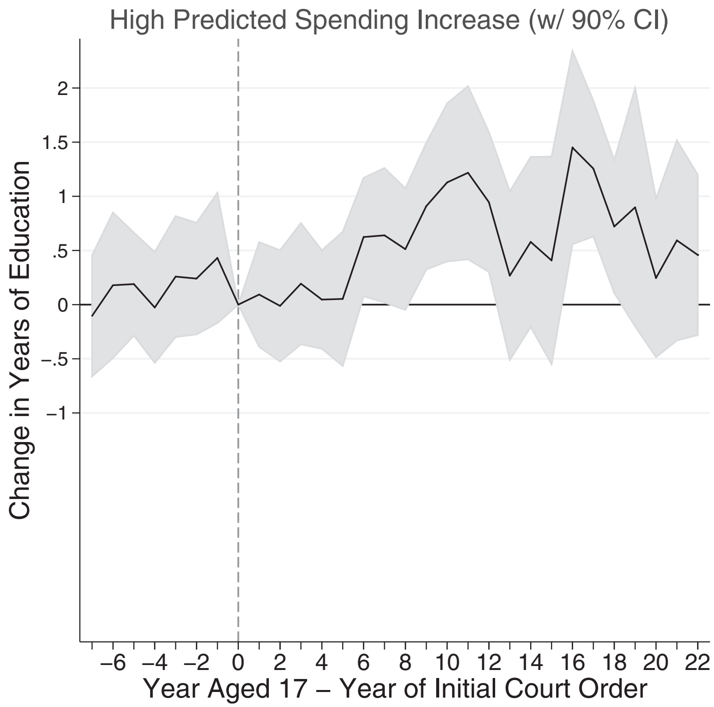


FIGURE III

Data: PSID geocode data (1968–2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011 ($N = 15{,}353$ individuals from 1,409 school districts [1,031 child counties, 50 states]). Sampling weights are used so that the results are nationally representative.

Models: The event study plot is based on indicator variables for the number of school-age years of exposure to a court-ordered SFR interacted with whether the district is predicted to experience a spending increase due to reforms ($Spend_{d\cdot} > 0$) or not. Results are based on nonparametric event study models that include school district fixed effects, race × census division × birth cohort fixed effects, and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race) each interacted with linear cohort trends.

from 0 at the 10 percent level. Both the patterns in timing and dosage support the hypothesis that policy-induced increases in school spending led to significant increases in educational attainment. We present very similar event study figures for the probability of high school graduation in Online Appendix J.

Having established visually that there are significant improvements in long-run educational attainment associated with policy-induced spending increases for exposed cohorts, we now quantify the causal relationship between actual school spending and longer-run educational attainment. For this we turn to the instrumental variable (IV) models that use the event study patterns to predict changes in childhood exposure to per pupil spending. Putting all the variation together, the 2SLS/IV models provide a direct estimate of the effect of school spending on adult outcomes and allow for tests of statistical significance.

The 2SLS/IV estimated effects of spending on educational attainment are presented in Table III. The explanatory variable of interest is the natural log of average per pupil spending during an individual's school years. The interpretation of a 0.10 and 0.20 change in this variable is the effect of increasing school spending by 10 percent and 20 percent throughout all 12 of an individual's school-age years, respectively. The excluded instruments for this spending variable are the number of school-age years of exposure to reforms ($Exp_{(t_{idb}-T_d^*)}$) and its interaction with indicator variables denoting the district's quartile in the distribution of predicted dosage ($Q_{Spend,d}$).To assuage any concerns about the construction of our $Spend_d$ variable, we also present results where our excluded instruments are the number of school-age years of exposure to reforms and its interaction with indicator variables denoting the district's quartile in the respective state distribution of per pupil spending in 1972 ($Q_{ppe1972,d}$). The first-stage F-statistic is greater than 10 in all models. For comparison purposes, we also present estimates from OLS regression models that do not account for the possible endogeneity of school spending.

Column (3) of Table III presents the 2SLS/IV regression results based on variation presented in Figures II and III for all children. The 2SLS estimates indicate that increasing per pupil spending by 10% in all 12 school-age years increases educational attainment by 0.31 years on average among all children. To put this effect size into perspective, a 13% spending increase is roughly the increase in spending experienced by cohorts that were five years old at the time of the initial court order in districts

TABLE III

OLS versus 2SLS Estimates of Court-Ordered School Finance Reform Induced Effects of Per Pupil Spending on Educational Attainment: By Childhood Poverty Status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable:							
	Years of Education				Prob(High School Graduate)			
	OLS	2SLS 1	2SLS 2		OLS	2SLS 1	2SLS 2	
Ln(PPE_d)$_{(\text{age 5-17})}$	−0.0763	3.1600***	3.1488***		0.0216	0.5910**	0.7053***	
	(0.1474)	(1.1327)	(0.7906)		(0.0278)	(0.2858)	(0.1436)	
Ln(PPE_d)$_{(\text{age 5-17})}$ × low income				4.5899***				0.9878***
				(1.2072)				(0.2744)
Ln(PPE_d)$_{(\text{age 5-17})}$ × Nonpoor				0.7156				0.2470
				(1.3193)				(0.1757)
Number of individuals	15,353	15,353	15,353	15,353	15,353	15,353	15,353	15,353
Number of childhood families	4,586	4,586	4,586	4,586	4,586	4,586	4,586	4,586
First-stage F-statistic	N/A	15.62	20.25	20.25	N/A	15.62	20.25	20.25

Notes. Robust standard errors in parentheses (clustered at school district level). *** $p < .01$, ** $p < .05$, * $p < .10$. Data: PSID geocode data (1968–2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative. Models: the key treatment variable, ln(PPE_d)$_{(\text{age 5-17})}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. 2SLS 1: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (quartile of the district in the state distribution of per pupil school spending in 1972). 2SLS 2: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$). Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race (proxy for segregationist preferences)) each interacted with linear cohort trends. In models by childhood poverty status, all variables are interacted with childhood poverty status.

with predicted spending increases. We also present 2SLS/IV estimates using the quartile of the district in the state distribution of per pupil spending in 1972 as our district-specific predictor of dosage (column (2)). Using this alternate instrument yields a point estimate on years of completed education of 3.16. This estimate is almost identical to that obtained using our more refined measure of dosage. However, the latter is less precise; the standard error of the point estimate is more than 40% larger than those obtained using our more refined measure of dosage, and the first-stage F-statistic is about 25% smaller. This underscores the efficiency gains from using more information about reform type and prereform district income levels in predicting ex ante reform-induced changes in spending.

Because residential mobility across counties and private school attendance are more common among affluent families than in low-income families, one might expect larger effects among low-income children.[24] Furthermore, prior research has shown that children from low-income families are more sensitive to certain school-related interventions than children from more advantaged backgrounds (e.g., Krueger and Whitmore 2001). Accordingly, we test for differential effects of school spending by childhood family income in column (4). The results reveal much larger effects for low-income children. For children from low-income families, increasing per pupil spending by 10% in all 12 school-age years increases educational attainment by 0.46 years (p-value $< .01$). In contrast, for nonpoor children, a 10% increase in per pupil spending throughout one's school-age years increases educational attainment by only 0.071 years, and this estimate is not statistically significant. To put these educational attainment estimates in perspective, the gap in completed years of education between children from low-income and nonpoor families is one full year (the average difference in childhood family income across these groups is about $40,000). Thus, the estimated effect of a 21.7% increase in per pupil spending throughout all 12 school-age years for low-income children is large enough to eliminate the educational attainment gap between children from low-income and nonpoor families. This

24. Prior research shows that although residential instability is greater for poor families, poor families are far less likely to move to better neighborhoods and are less responsive to policy changes due to the greater residential location constraints they face (Johnson 2008; Kunz et al. (2003); Jackson et al. 2007).

relatively large increase is within the range of the variation induced by SFRs and corresponds to the increase in spending for cohorts who were born the year of the initial court order in districts with larger spending increases (i.e., districts with $Spend_d > 0$). In relation to recent spending levels (the average for 2011 was $12,600 per pupil in 2013 dollars), this would correspond to increasing per pupil spending permanently by roughly $2,900 per student in 2015 dollars.

To examine the margin of educational attainment affected, columns (7) and (8) present the 2SLS regression estimates for the likelihood of high school graduation using the preferred, more refined instruments. Overall, the 2SLS estimate indicates that increasing per pupil spending by 10% in all 12 school-age years increases the probability of high school graduation by 7 percentage points. Results using the simple, but coarser instrument (column (6)) are slightly smaller and suggest that increasing per pupil spending by 10% in all school-age years increases the probability of high school graduation by 5.9 percentage points overall. Looking by childhood poverty status, the preferred 2SLS estimate indicates that increasing per pupil spending by 10% in all school-age years increases the probability of high school graduation by 9.8 percentage points (p-value $< .01$) for low-income children but only 2.4 percentage points (not statistically significant) for nonpoor children. The 95% confidence interval for the effect of a 10% increase for low-income children is between 2.5 and 15.2 percentage points. The high school graduation rates for low-income and nonpoor children were 79% and 92%, respectively. Accordingly, among low-income children, increasing per pupil school spending by 10% over the entire schooling career increases the likelihood of graduating from high school by between 5.6% and 19.3%. These results indicate large positive effects for low-income children and suggest small positive effects for more affluent children.

To put these estimates in perspective, attending Head Start or the Perry preschool program increased high school completion by 8.5 and 14 percentage points, respectively (Carneiro and Heckman 2003; Deming 2009). Also, Barrow, Claessens, and Schanzenbach (2013) and Schwartz, Stiefel, and Wiswall (2013) find that attending small schools increases graduation rates for low-income children by 16 to 18 percentage points. Accordingly, our effects on educational attainment, although large, are somewhat smaller than those of some very successful interventions. In sum, both the event study and 2SLS/IV models reveal that exogenous increases in school spending (caused by SFRs) led to substantial improvements in educational outcomes of affected children. Regression results indicate that there are much larger effects of school spending on educational attainment for children from low-income families.

V.B. Labor Market Outcomes, Adult Family Income, and Poverty Status

The next series of results reveal economically meaningful effects of school spending on low-income children's subsequent adult economic status and labor market outcomes, using the same model specifications. It is important to note that our models that analyze economic outcomes (such as wages and annual family income) use all available person-year observations for ages 20–45 and control for a cubic in age to avoid confounding life cycle and birth cohort effects. As with the educational outcomes, we present event study graphs of court-mandated SFR effects on adult economic outcomes (ages 20–45) by predicted dosage and then present the regression results for each outcome.

Our estimated spending effects on economic outcomes mirror those on educational attainment. We first discuss the reform-induced spending effects on adult wages for the full sample (Figure IV). Overall, one can see clear patterns of improved economic outcomes for exposed cohorts from districts with larger predicted spending increases. Among unexposed cohorts, we find no discernible trending in wages, and the pattern of prereform event study year estimates is very similar for those from both districts with low and high predicted dosage and those from nonreform districts. Placebo tests using spending during non-school-age years presented below support this conclusion. As with the education results, these event study graphs capture the reduced-form effects of predicted spending increases, not effects of actual spending increases (estimates of actual spending are provided in the 2SLS regression results). Among exposed cohorts, those cohorts with more years of exposure to larger predicted spending increases (Bottom panel) have higher wages than unexposed cohorts and cohorts with fewer years of exposure. Indeed, after five years of exposure, one can reject the null hypothesis that most of the event study years are different from that of no exposure at the 10% level. Importantly, we find no

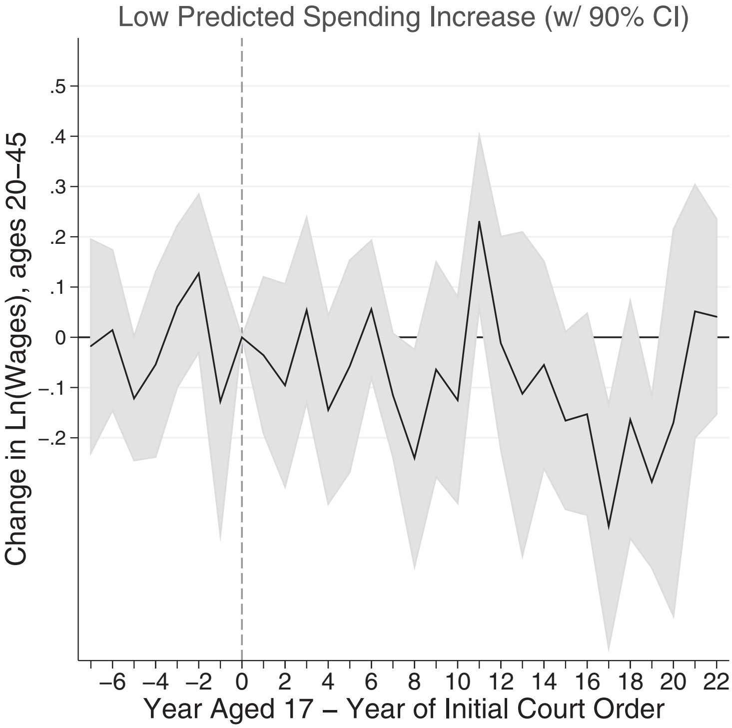


FIGURE IV

Effect of Court-Ordered School Finance Reform on ln(Wage)

High predicted spending increase refers to districts in reform states with $Spend_{d\cdot} > 0$ and low predicted spending increase refers to districts in reform states with $Spend_{d\cdot} \leq 0$. Roughly two thirds of districts in reform states had $Spend_{d\cdot} > 0$.

Continued.

systematic statistically significant effects on adult wages for exposed cohorts from districts with $Spend_d \leq 0$ that saw little to no change in actual school-age per pupil spending.[25] These results reinforce a consistent pattern, and provide compelling evidence that the effect of the reforms on outcomes operate through the effects on spending (as opposed to other possible factors). Very similar event study figures for adult family income are in Online Appendix J.

25. Only 3 of the 22 postreform event study years are statistically significantly different from that of no exposure, and they do not all have the same sign.

Continued

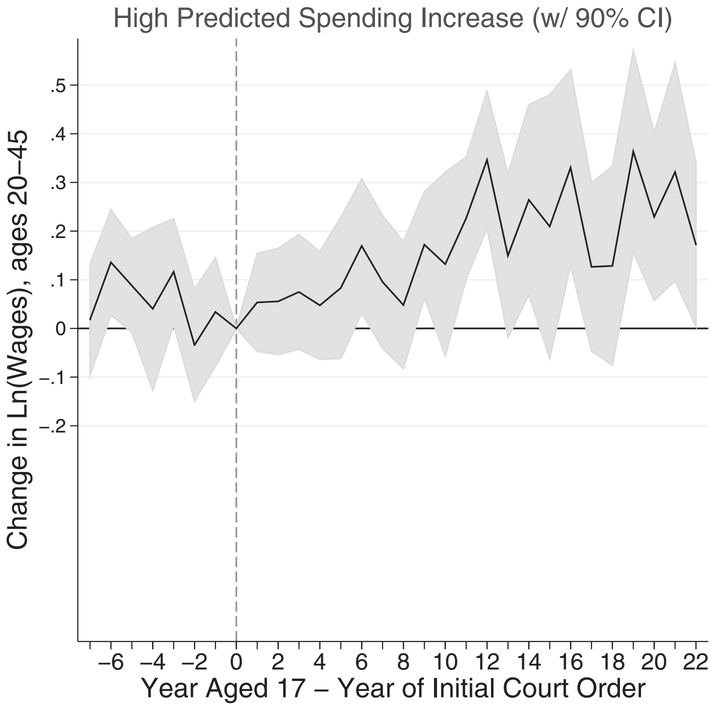


FIGURE IV

Data: PSID geocode data (1968–2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011 ($N = 15{,}353$ individuals from 1,409 school districts [1,031 child counties, 50 states]). Sampling weights are used so that the results are nationally representative.

Models: The event study plot is based on indicator variables for the number of school-age years of exposure to a cour-ordered SFR interacted with whether the district is predicted to experience a spending increase due to reforms ($Spend_{d.} > 0$) or not. Results are based on nonparametric event study models that include school district fixed effects, race × census division × birth cohort fixed effects, and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race) each interacted with linear cohort trends.

The 2SLS/IV estimates for all the adult economic outcomes are presented for all children and separately by childhood economic status in Table IV. Using the quartile of the district in the state distribution of per pupil spending in 1972 interacted with years of exposure as the excluded instruments, the estimate for the full sample is 0.7076 (p-value < .01). The preferred model that uses the quantile of $Spend_d$ interacted with years of exposure as the excluded instruments yields a similar point estimate of 0.774 (p-value < .01). As with the education results, the point estimate is almost identical but the standard errors are almost 40% larger using the simple instrument as opposed to the preferred more refined instrument. This emphasizes the robustness of our results. Our preferred 2SLS/IV model implies that on average, increasing per pupil spending by 10% in all school-age years increases adult wages by 7.74%. Consistent with the larger effects on educational attainment for children from low-income families, the adult wage effects are more pronounced for children from low-income families. As shown in column (4), the preferred 2SLS/IV estimates reveal that for children from low-income families, increasing per pupil spending by 10% in all school-age years increases adult wages by about 9.6% (p-value < .01). This implies an elasticity of wages with respect to per pupil spending close to 1 for children from low-income families. However, the 95% confidence interval of this estimate supports a range of elasticities between a modest 0.37 and a sizable 1.54. In contrast, the 2SLS estimate for children from nonpoor families is smaller and statistically insignificant. It is worth noting that the point estimate implies that for nonpoor children, increasing per pupil spending by 10% in all 12 school-age years increases adult wages by 5.5%. Although this effect is not statistically significant, the effect is economically important and is suggestive of benefits for all children, with larger effects for those from low-income families.

Although some of these wage effects will be due to increased years of schooling (for those induced to stay in school longer), the effect of improved school quality on those who do not change their school-going behaviors will be reflected in their wages but not their years of schooling.[26] To put our estimates in context, it is

26. Recent studies find that improvement in instruction are reflected in improved outcomes above and beyond their effects on years of schooling. Goodman (2012) finds that an addition year of math coursework in high school increases black males' earnings by 5–9%, conditional on overall years of schooling.

TABLE IV

OLS VERSUSS 2SLS ESTIMATES OF COURT-ORDERED SCHOOL FINANCE REFORM INDUCED EFFECTS OF PER PUPIL SPENDING ON ADULT WAGES AND FAMILY INCOME: BY CHILDHOOD POVERTY STATUS (ALL ADULT OUTCOMES ARE MEASURED BETWEEN AGES 20–45)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable:							
	Ln(Wage), Ages 20–45				Ln(Family Income), Ages 20–45			
	OLS	2SLS 1	2SLS 2		OLS	2SLS 1	2SLS 2	
Ln(PPE_d)$_{(\text{age 5-17})}$	−0.0480	0.7076***	0.7743***		0.0128	0.8705***	0.9819***	
	(0.0654)	(0.2679)	(0.1959)		(0.0617)	(0.3348)	(0.2849)	
Ln(PPE_d)$_{(\text{age 5-17})}$ × Low income				0.9598***				1.7146***
				(0.3003)				(0.3585)
Ln(PPE_d)$_{(\text{age 5-17})}$ × Nonpoor				0.5525				0.2021
				(0.4461)				(0.4223)
Number of person-year obs.	106,545	106,545	106,545	106,545	151,349	151,349	151,349	151,349
Number of individuals	13,183	13,183	13,183	13,183	14,730	14,730	14,730	14,730
Number of childhood families	4,454	4,454	4,454	4,454	4,588	4,588	4,588	4,588
First-stage F-statistic	N/A	15.62	20.25	20.25	N/A	15.62	20.25	20.25

Notes. Robust standard errors in parentheses (clustered at school district level). *** $p < .01$, ** $p < .05$, * $p < .10$. Data: PSID geocode data (1968–2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative. Models: the key treatment variable, ln(PPE_d)$_{(\text{age 5-17})}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. 2SLS 1: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (quartile of the district in the state distribution of per pupil school spending in 1972). 2SLS 2: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (quartile of the district in the distribution of $Spend_d$). Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race (proxy for segregationist preferences)) each interacted with linear cohort trends. In models by childhood poverty status, all variables are interacted with childhood poverty status.

helpful to determine how much of the estimated effect on wages can plausibly be attributed only to increases in years of schooling. Tables III and IV indicate that increasing school spending by 10% for all of a low-income child's school-age years will increase their years of schooling by 0.46 years and their adult wages by 9.6%. Recent credible estimates of the returns to an additional year of schooling indicate returns between 9% and 28%,[27] so that wage effects between 3.87% and 12.04% can be expected only through a years of schooling effect. The actual increase in wages of 9.6% is well within this range. Because years of education may only capture some of the effect on wages, our estimates are consistent with effect sizes suggested by the existing literature.

Note that although we look at individuals in their forties, recent studies of interventions on earnings tend to look at individuals in their twenties (e.g., Chetty et al. 2014). The estimated effects on wages for those in their twenties likely understates the effect on permanent income. To assess the importance of this, we estimate the effect of school spending interacted with a cubic in an individual's age. The implied age profile of the school spending effects on wages are presented in Online Appendix J. One can reject the null of no age profile at the 5% level. The results imply that the increase in wages that result from a 10% increase in school spending throughout the school-age years is 2.8% at age 20, about 8% during one's thirties, and 13.4% at age 45. Another important aspect of our data is that we observe the same individuals in multiple years rather than at one point in time (as in the CPS or census). As such, individuals with low labor market attachment (who might be highly responsive to improvements in school quality) who might not have earnings in any given year can be observed with earnings in the PSID at some point over the panel, which minimizes potential sample selection bias. To show

Also, Fredriksson et al. (2012) find that the effects of class size on earnings are much larger than imputed effects based on increases in years of education.

27. Older estimates range from between 7.2% (Angrist and Krueger 1991, using only males in the 1980 census) to 16% (Ashenfelter and Krueger 1994, using both males and females in 1990 CPS). From Katz and Autor (1999), we know that the wage premium has been increasing over time, such that estimates from the 1980s and 1990s are likely to provide a lower bound to what one might expect for a cohort of workers in 2010. Jepsen, Troske, and Cooms (2014) use recent data and find wage returns due to an additional year of community college enrollment as high as 28% for women and 14% for men.

this, we estimate wage effects using only wage outcomes from the 2001 or 2011 waves (Online Appendix J). Using a single year of wage data yields point estimates between half and two-thirds as large as those using all years, and yields standard errors four times as large. Haider and Solon (2006) show that using a single year of data, and data at young ages, will lead one to understate effects on earnings. This appears to be an important reason why our wage estimates are somewhat larger than others in the literature.

We also estimate effects on family income. The results from the 2SLS/IV models for adult family income are similar to those of other outcomes. As shown in column (7), the results indicate that on average, increasing per pupil spending by 10% in all 12 school-age years increases family income by 9.8% (p-value $< .01$). As with the other outcomes, the average benefits overall are driven by large effects for children from low-income families. Column (8) shows that for children from low-income families, increasing per pupil spending by 10% in all 12 school-age years increases family income by 17.1%, and this estimate is significant at the 1% level. For children from low-income families, the 95% confidence interval for a 10% spending increase is between 10.1% and 24.2%. For children from nonpoor families, the estimated effect is small and not statistically significant at the 10% level. The effects on family income reflect (i) increases in own income, (ii) increases in other income due to increases in the likelihood of being married (i.e., there are more potential earners), and (iii) increase in the income of one's family members (which is likely if persons marry individuals who were also affected by spending increases). Consistent with the effects on family income reflecting in part a family composition effect, we find that among low-income children, a 10% spending increase is associated with a 10 percentage point increased likelihood of currently being married and never previously divorced (not shown). There is no effect on the probability of ever being married, so this appears to reflect a marital stability effect.

Our final measure of overall economic well-being is the annual incidence of adult poverty. Because this is an undesirable outcome, estimates should be interpreted such that lower numbers are better. The event study is presented in Figure V. As with the other outcomes, there is evidence of a causal effect of school spending on outcomes. There is no prereform trending in outcomes across unexposed cohorts. However, exposed cohorts from districts with larger predicted spending increases have

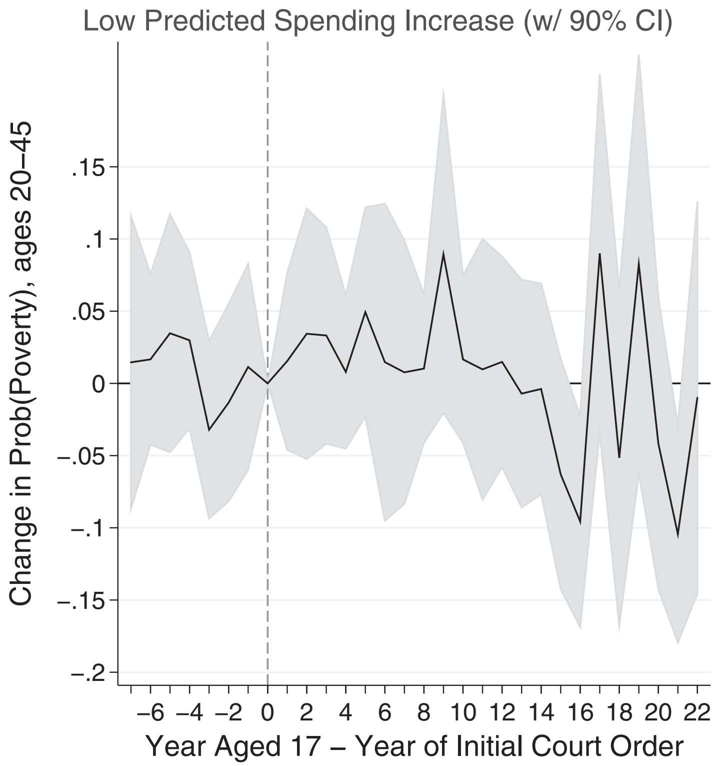


FIGURE V

Effect of Court-Ordered School Finance Reform on Annual incidence of Adult Poverty

High predicted spending increase refers to districts in reform states with $Spend_{d\cdot} > 0$ and low predicted spending increase refers to districts in reform states with $Spend_{d\cdot} \leq 0$. Roughly two-thirds of districts in reform states had $Spend_{d\cdot} > 0$.

Continued.

steady declines in the annual incidence of adult poverty that become more pronounced with years of exposure (Bottom). In contrast, the event study for districts with low predicted spending increases (Top) shows no systematic change in outcomes across cohorts. The 2SLS/IV results are presented in Table V and mirror the findings from the event study models. In the preferred model (column (3)), the 2SLS/IV estimate for all children indicates that increasing per pupil spending by 10% in all 12 school-age years reduces the annual incidence of poverty in adulthood by 2.67

Continued

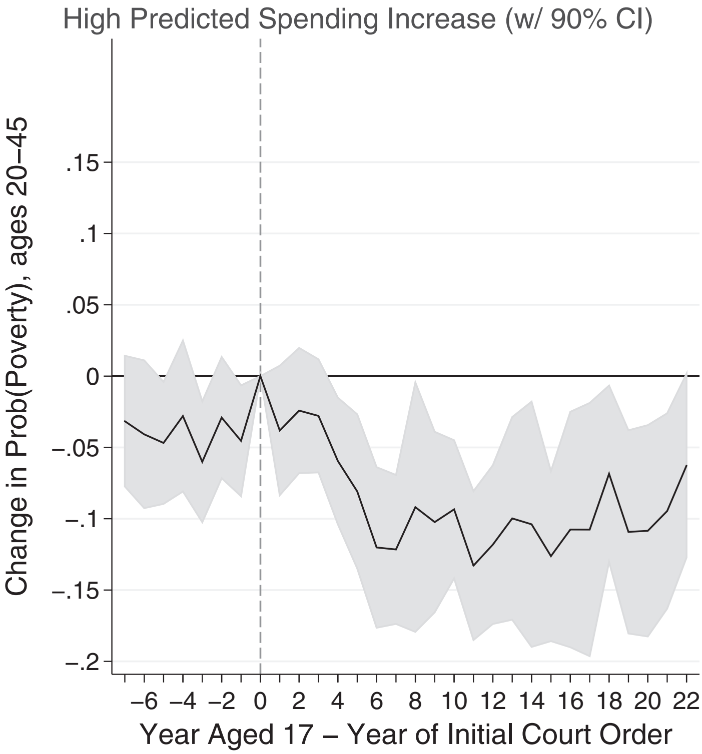


FIGURE V

Data: PSID geocode data (1968–2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011 ($N = 15{,}353$ individuals from 1,409 school districts [1,031 child counties, 50 states]). Sampling weights are used so that the results are nationally representative.

Models: The event study plot is based on indicator variables for the number of school-age years of exposure to a court-ordered SFR interacted with whether the district is predicted to experience a spending increase due to reforms ($Spend_{d\cdot} > 0$) or not. Results are based on nonparametric event study models that include school district fixed effects, race × census division × birth cohort fixed effects, and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race) each interacted with linear cohort trends.

TABLE V

OLS versus 2SLS Estimates of Court-Ordered School Finance Reform Induced Effects of Per Pupil Spending on Adult Poverty Status: By Childhood Poverty Status (All Adult Outcomes Are Measured between Ages 20–45)

	(1)	(2)	(3)	(4)
	Dependent Variable: Prob(Poverty), Ages 20-45			
	OLS	2SLS 1	2SLS 2	
$\text{Ln}(PPE_d)_{(\text{age 5-17})}$	−0.0045	−0.3228***	−0.2678***	
	(0.0124)	(0.0763)	(0.0710)	
$\text{Ln}(PPE_d)_{(\text{age 5-17})} \times$ Low income				−0.6132***
				(0.1242)
$\text{Ln}(PPE_d)_{(\text{age 5-17})} \times$ Nonpoor				0.0385
				(0.0850)
Number of person-year obs.	151,756	151,756	151,756	151,756
Number of individuals	14,737	14,737	14,737	14,737
Number of childhood families	4,588	4,588	4,588	4,588
First-stage F-statistic	N/A	15.62	20.25	20.25

Notes. Robust standard errors in parentheses (clustered at school district level). *** $p < .01$, ** $p < .05$, * $p < .10$. Data: PSID geocode data (1968–2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative. Models: The key treatment variable, $\ln(PPE_d)_{(\text{age 5-17})}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. 2SLS 1: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (quartile of the district in the state distribution of per pupil school spending in 1972). 2SLS 2: the excluded instruments from the second stage are (number of years of exposure to a court-ordered SFR) and (number of years of exposure to a court-ordered SFR) × (quartile of the district in the distribution of $Spend_d$). Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at child's birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race (proxy for segregationist preferences)) each interacted with linear cohort trends. In models by childhood poverty status, all variables are interacted with childhood poverty status.

percentage points. Results by childhood family income reveal that this average effect is driven entirely by children from low-income families (column (4)). The 2SLS/IV estimate for children from low-income families indicates that increasing per pupil spending by 10% in all school-age years reduces the annual incidence of poverty in adulthood by 6.1 percentage points. This estimated effect is statistically significant at the 1% level and the 95% confidence interval is between 3.7 and 8.56 percentage points. The effect for children from nonpoor families is small and not statistically significantly different from zero.

In summary, increases in school spending led to increases in adult economic attainment that rose in line with their educational improvements (likely reflecting a combination of improvements in both the quantity and quality of education). These average effects were driven largely by sizable improvements for children from low-income families. Taken together, the event study graphs and the IV regression estimates based on exogenous changes in school spending show that increased reform-induced spending had meaningful causal effects on educational attainment, adult wages, family income, and adult poverty status. We now present a series of robustness tests and discuss the findings in the context of prior studies in the literature.

V.C. Robustness Checks

1. Falsification Tests. If the effects documented are causal effects of school spending, effects should be present for spending changes that occur during school-age years with no corresponding significant effects for spending that occur during non-school-age years. As such, as a placebo falsification test we test whether reform-induced spending changes that occur after individuals should have left school (between the ages of 20 and 24) affects outcomes. This falsification exercise is detailed in Online Appendix K. For all outcomes, the point estimates are small, and there are no statistically significant effects of reform-induced spending that occurred when individuals were between the ages of 20 and 24. This supports a causal interpretation of our estimates.

2. Addressing Endogenous Residential Mobility. One may worry that our results are biased by endogenous residential mobility. To address potential bias, we reestimated all models limiting the analysis sample to those who lived at their (earliest) childhood residence prior to the enactment of initial court orders in their respective state. The results are presented in Online Appendix L. We find nearly identical results as those in the full sample. This indicates that endogenous residential mobility is not an important source of bias in our analysis.

3. Addressing Bias Due to Recent Reforms. Even though we are careful to control for several policies, we explored whether our results are affected by the more recent policy reforms that started in the late 1980s (such as charter schools and test-based accountability). To test for this, we estimate separate school spending effects for those born between 1970 and 1985 and those born between 1955 and 1969 (see Online Appendix M). If our effects are driven by other recent reforms, there should be no effect for the older cohorts, and the effect for more recent cohorts should be statistically significantly different from that of the older cohorts. On the contrary, there is no statistically significant difference between the marginal effects for the older versus more recent cohorts. This suggests little to no bias due to other more recent reforms.

4. Validating Using Other Data. To ensure that our estimated patterns generalize to all school districts (not just those in the PSID), we replicated the analyses for high school graduation using aggregate high school graduation rates from the Common Core Data (CCD) for all school districts in the United States for available years 1987–2010 with the preferred research design (see Online Appendix N). School spending effects on the number of graduates per eighth-grader are on a similar order of magnitude as the graduation rate estimates from the PSID. We also employ census and American Community Survey (ACS) data for the same birth cohorts and ages as those covered in the PSID. Using state-level variation in spending, we find that increases in per pupil spending lead to increases in years of education and earnings that are in line with the estimates from the PSID.

V.D. The Importance of Using Exogenous Variation

As mentioned previously, merely correlating changes in spending with changes in outcomes could yield biased results. To gauge the extent to which this matters, we compare our estimated naive OLS regression to the 2SLS regression estimates. For all outcomes and subsamples, the OLS estimates are orders of magnitudes smaller than the 2SLS/IV estimates. Looking at the education outcomes in the PSID sample (Table III, columns (1) and (5)), OLS estimates show no statistically significant relationship between school spending and outcomes. As further evidence of no effects using observational variation, both OLS point estimates are small and the point estimates for high school graduation and years of education have opposite signs. The OLS estimates for the economic outcomes show a similar pattern in Tables IV and V. The naively estimated school-spending effects are close to zero and go in opposite directions—indicating no relationship between potentially endogenous variation in school spending and adult outcomes, despite large effects of exogenous spending increases on adult outcomes. To ensure that this is not an artifact of the PSID data, we replicate this same pattern in the CCD data and the census data of small estimated relationships using all variation in school spending and large positive effects using the reform-induced variation in school spending (Online Appendix N).

The stark contrast between the OLS and the 2SLS estimates across all three data sets provides an explanation for why these estimates might differ from other influential studies (e.g., Coleman et al. 1966; Betts 1995; Hanushek 1996; Grogger 1996). Prior studies that relied on actual variation in spending may have produced modest effects of school spending due to unresolved endogeneity biases. Indeed, in Table II, we show that noninstrumented within-district increases in school spending are significantly related to increases in childhood family/community socioeconomic disadvantage, whereas instrumented school spending is not. This suggests that OLS estimates are likely biased against finding a positive school spending effect and makes clear the need for exogenous variation in school spending. However, Table II also provides another possible reason for the difference in findings: noninstrumented school spending is unrelated to better school inputs while instrumented school spending is. As such, another potential explanation for our finding large school spending effects is that how the money is spent matters a lot, and that exogenous increases in school spending are more closely tied to productive inputs than endogenous increases in school spending.[28] Given that money per se will not improve student outcomes (for example, using the funds to pay for lavish

28. This finding prompts the question of why school districts are more likely to reduce class sizes and improve other inputs with an exogenous windfall of school spending than endogenous changes in school spending. Though investigating this is outside the scope of this project, one possible explanation with anecdotal support is that teachers' unions may be much more likely to demand higher salaries and smaller class sizes when they know that the district has recently received additional state funding. Indeed, teachers' unions in New Jersey and New York explicitly advise that members use information about state funding to gain leverage for

faculty retreats will likely not have a positive effect on student outcomes), understanding how the increased funding was spent is key to understanding why we find large spending effects where others do not. We explore these issues below.

VI. Exploring Mechanisms

To shed light on the mechanisms through which various types of education spending affects subsequent adult outcomes, we examine the effects of exogenous spending increases on spending for school support services, physical capital spending, and instructional spending. We also estimate effects on student-teacher ratios, student–guidance counselors ratios, teacher salaries, and the length of the school year (key input measures employed in the seminal literature on school quality). We employ data on the types of school spending (available for 1992–2010 from the CCD), student-staff ratios (available for 1987–2010 from the CCD and Office of Civil Rights), and information on teacher salaries and length of the school year (available approximately every three years for 1987–2010) from the School and Staffing Survey housed at Institute of Education Sciences. The earliest CCD data start in 1987, so we do not have detailed data for the same cohorts that are exposed to the early reforms in the PSID. However, an analysis of mechanisms for the more recent cohorts may be instructive.

To determine how each additional dollar associated with reforms was spent, we employ instrumental variables models similar to equations (1) and (2) where the main outcomes are the various school inputs. For ease of interpretation we present effects on the type of expenditure in levels. The interpretation of the estimate is the marginal propensity to spend (i.e., the increase in a particular type of spending associated with a $1 increase in total spending). For all other outcomes we use logs as in the rest of the article. The endogenous regressor is per pupil spending or log per pupil spending, and the excluded instrument is the number of years of exposure to reforms interacted with the district-specific spending increase. Results are presented in Table VI.

smaller classes and higher salaries. See http://finance.tc-library.org/Content.asp?uid=1305 and https://www.njsba.org/mo/labor_relations/home.php.

TABLE VI

EVIDENCE ON MECHANISMS (USING DATA FROM MULTIPLE SOURCES)

Dependent Variable	(1) Capital Spending Per Pupil	(2) Instructional Spending Per Pupil	(3) Services Spending Per Pupil	(4) Other Spending Per Pupil	(5) Students per Teacher	(6) Students per Guidance Counselor	(7) Students per Admin	(8) Length of School Year	(9) Teacher Base Salary
Source of Data	INDFIN	Common Core Data			Common Core Data			School and Staffing Survey	
Years Outcome Available	Annually 1972 - 2010	Annually 1987–2010			Annually 1987–2010			1987, 1990, 1993, 1999, 2003, 2007, 2011	
Spending	0.106***	0.668***	0.408***	−0.0598***					
	(0.0268)	(0.0665)	(0.0457)	(0.0149)					
Ln(per pupil spending)					−7.017***	−513.2***	−275.3***	13.56**	10,764***
					(0.750)	(104.6)	(74.71)	(5.343)	(2,505)
Observations	498,708	260,290	260,290	260,290	293,544	200,474	222,951	23,807	23,587
Number of districts	17,849	16,094	16,094	16,094	16,172	13,802	15,012	7,011	6,938
Mean of dep. var.	828.4	6,213	3,508	428.6	12.14	411.3	243.3	179	26,331

Notes. Robust standard errors in parentheses (clustered at school district level). *** $p < .01$, ** $p < .05$, * $p < .1$. Models: Results are based on 2SLS/IV models that include school district fixed effects; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) each interacted with linear time trends. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) $\times$ (the quartile of the district in the distribution of $Spend_d$). All models yield first-stage F-statistics greater than 20.

When a district increases school spending by \$100 due to reforms, spending on capital increases by \$10.60, spending on instruction increases by \$66.80, and spending on support services increases by \$40.80 on average. Instructional spending makes up about 60% of all spending, and it accounts for about two thirds of the marginal increase. Also, spending on support services makes up about 32% of all spending, and it accounts for about 40% of the marginal increase. This suggests that on the margin, exogenous increases in school spending may be somewhat more likely to go to instruction and support services than other spending increases. To account for this increase, districts that experience increases in total spending tend to see declines in other spending (noninstructional, nonsupport services, noncapital spending). The increases for instruction and support services (which includes expenditures to hire more teachers and/or increase teacher salary along with funds to hire more guidance counselors and social workers) are consistent with the large, positive effects for those from low-income families.

Prior research has emphasized that an important determinant of how much students learn is teacher quality; teachers' salaries represent the largest single cost in K–12 education and may exert a direct effect on the ability to attract and retain a high-quality teaching workforce. The largest share of school districts' spending (annual operating budgets—instructional expenditures) is composed of two components: (i) the number of teachers hired, which governs the teacher-student ratio; and (ii) the salary schedule (by qualifications—experience and educational background credentials). Accordingly, we next separately estimate effects on average teacher salaries and student-staff ratios. For these models, the endogenous regressor is the natural log of school spending. Districts that increased spending due to reforms see reductions in student-teacher ratios. This has been found to benefit students in general, with larger effects for children from disadvantaged backgrounds (e.g., Krueger and Whitmore, 2001; Bloom and Unterman, 2013). To show that our effects on student-teacher ratios track the increases in school spending, we linked the school spending data to the PSID sample and augmented these data with student-teacher data at the district level during 1968–1977 from the Office of Civil Rights. We then estimated our event study models on student-teacher ratios. The results are presented in Figure VI. The results clearly show that there were no preexisting time trends in student-

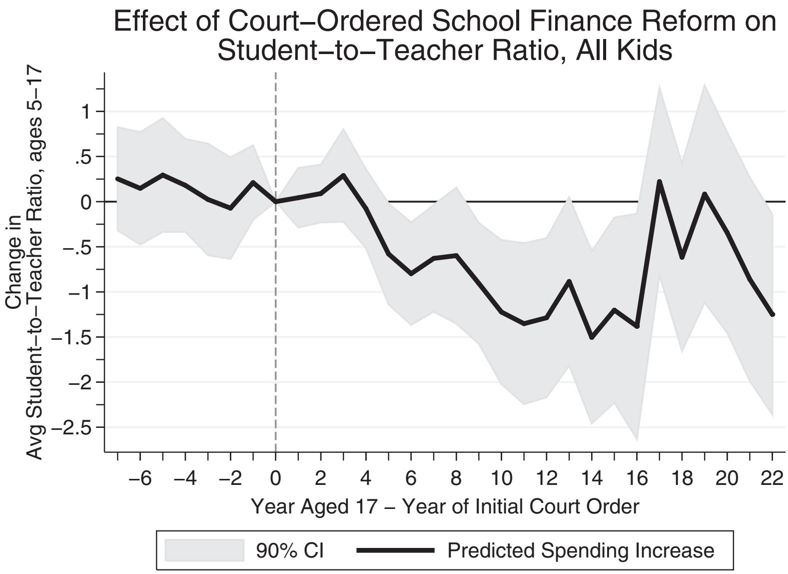


FIGURE VI

Effect of Predicted Reform Induced Spending Changes Interacted with Time Relative to the First Court-Ordered Reform on Student-Teacher Ratios

Data: PSID geocode data (1968–2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955–1985, followed into adulthood through 2011 ($N = 15{,}353$ individuals from 1,409 school districts [1,031 child counties, 50 states]). Sampling weights are used so that the results are nationally representative.

Models: The event study plot is based on indicator variables for the number of school-age years of exposure to a court-ordered SFR interacted whether the district is predicted to experience a spending increase due to reforms ($Spend_{d\cdot} > 0$) or not. Results are based on nonparametric event study models that include school district fixed effects, race × census division × birth cohort fixed effects, and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age four), food stamps, Medicaid, AFDC, UI, Title I (average during childhood years), timing of state-funded kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 presidential election * race) each interacted with linear cohort trends.

teacher ratios, the decreases in student-teacher ratios coincide with the passage of school finance reforms, and the reduction in student-teacher closely tracks the reform-induced increases in spending.

We also find that schools in these districts have fewer students per counselor and fewer students per administrator, which have also been found to improve student outcomes (e.g., Carell and Carell 2006; Reback 2010). In addition to using student-teacher ratios, Card and Krueger (1992) and Loeb and Bound (1996) proxy for school quality with the length of the school year and teacher salaries. We also analyze effects on these measures. The 2SLS estimates indicate that a 10% increase in school spending is associated with a 5.7% reduction in the student-teacher ratio (p-value $< .01$), 1.36 more school days (p-value $< .01$), and a 4% increase in base teacher salaries (p-value $< .01$). Insofar as these mechanisms are partly responsible for the improved student outcomes, these findings stand in stark contrast to studies finding little effect of these measures on student outcomes for cohorts educated after 1950 (Betts 1995, 1996; Hanushek 2001). The results are in line with studies on recent cohorts that use randomized and quasi-random variation in school inputs (e.g., Fredriksson et al. 2012; Chetty et al. 2013;), further underscoring the limitations of using observational variation for these important questions.[29]

Although there may be other mechanisms through which increased school spending improves student outcomes, the results suggest that the positive effects are driven, at least in part, by some combination of reductions in class size, having more adults per student in schools, increases in instructional time, and increases in teacher salary that may have helped attract and retain a more highly qualified teaching workforce.[30]

29. These studies are based on exogenous variation in particular school inputs at the individual or classroom level holding all other inputs fixed. Here we study district-wide changes in spending.

30. Class sizes are roughly 1.4 times larger than student-teacher ratios, so that our estimates imply a class size reduction of 0.98 students for a 10% spending increase. Fredriksson et al. (2012) find that wages are 0.0063 higher for three years with one less student in class. If we multiply this by our reduction of 0.98 students for 12 years, this implies a wage increase of 2.5%. As such, about one-third of our wage effect can plausibly be attributed to class size.

VII. Discussion and Conclusions

Previous national studies correlated observed school resources with student outcomes and found little association for those born after 1950 (e.g., Coleman et al. 1966; Hanushek 1986; Betts 1995; Grogger 1996). This study builds and improves on previous work by using nationally representative, individual-level panel data from birth to adulthood (matched with school spending and reform data) and quasi-experimental methods to estimate credible causal relationships. We investigate the causal effect of exogenous school spending increases (induced by the passage of SFRs) on educational attainment and (eventual) labor market success. For children from low-income families, increasing per pupil spending yields large improvements in educational attainment, wages, family income, and reductions in the annual incidence of adult poverty. All of these effects are statistically significant and are robust to a rich set of controls for confounding policies and trends. For children from nonpoor families, we find smaller effects of increased school spending on subsequent educational attainment and family income in adulthood. The results make important contributions to the human capital literature and highlight how improved access to school resources can profoundly shape the life outcomes of economically disadvantaged children, and thereby significantly reduce the intergenerational transmission of poverty.

To explore the potential mechanisms from which these spending effects arise, we documented that reform-induced school spending increases were associated with sizable improvements in measured school inputs, including reductions in student-teacher ratios, increases in teacher salaries, and longer school years.[31] These finding parallel those of Card and Krueger's influential 1992 study of males born between 1920 and 1949 and recent studies that link adult outcomes to quasi-experimental variation in school inputs (Fredriksson et al. 2012). The similarities suggest that money still matters, and so do school resources.

A suggestive benefit-cost analysis reveals that investments in school spending are worthwhile. Increasing spending by 10% for all school-age years increased wages by 7.7% each year (Table IV). Someone born in 1975 would start school around

31. These improvements also likely led to improvements in unobserved teacher quality (Jackson 2009, 2013).

1980 when average per pupil spending was $5,459 in 2013 dollars. A 10% increase for 12 years starting in 1980 is equal to $4,850 in present value (assuming a 6% discount rate). The median worker in 2013 earned $28,031, so a 7.2% increase in earnings for such a worker between ages 25 and 60 is worth just over $10,000 in present value. This implies a benefit-cost ratio of about 3 and an internal rate of return of roughly 10%. This internal rate of return is similar to those estimated for preschool programs (Deming 2009), smaller than estimates of the internal rates of return for class size reductions (Fredriksson et al. 2012), and larger than long-term returns to stocks. In sum, the estimated benefits to increased school spending are large enough to justify the increased spending under most reasonable benefit-cost calculations.

Given that school spending levels have risen significantly since the 1970s, our results might lead one to expect to have seen improved outcomes for children from low-income families, and indeed, other research suggests this occurred over the relevant time period. For example, Krueger (1998) documents test score increases over time, with large improvements for disadvantaged children from poor urban areas.[32] The CPS shows declining dropout rates since 1975 for those from the lowest income quartile (NCES 2012). Murnane (2013) finds that high school completion rates have been increasing since 1970 with larger increases for black and Hispanic students; Baum, Ma, and Payea (2013) find that postsecondary enrollment rates have been increasing since the 1980s, particularly for those from poor families. Our results suggest increased school spending may have played a key role.

Given that per pupil spending roughly doubled between 1970 and 2000, our point estimates might lead one to expect much greater convergence in outcomes across income groups. To help explain this, we point to studies documenting countervailing forces such as increased residential segregation by income (Reardon and Bischoff 2011; Watson 2009; Owens 2015),

32. Note, however, that Reardon (2013) finds that the gap between those at the 90th and 10th percentile of the income distribution (one of many measures of inequality) has been growing over time. He attributes this growth to improvement at the top of the income distribution rather than deterioration at the bottom. Also, his measure does not capture changes at other points in the income distribution. As such, the patterns documented in Reardon (2013) are not inconsistent with improved outcomes for the poor documented in Krueger (1998).

increases in single-parent families (Guryan, Hurst, and Kearney 2008; Waldfogel, Craigie, and Brooks-Gunn 2010), the crack epidemic (Evans, Garthwaite, and Moore 2012; Fryer et al. 2013), and mass incarceration (Raphael and Stoll 2009; Kearney et al. 2014). All of these forces tend to have large deleterious effects on those from low-income families. It is therefore likely that any positive school spending effects were offset by deteriorating conditions for low-income children in other dimensions. Aside from these countervailing forces, our evidence suggests that exogenous spending increases went toward more productive inputs than endogenous spending increases. Accordingly, our results predict that the effect of endogenous aggregate increases in school spending will be smaller than those implied by our estimates. Finally, we point out that we find that a 25% increase in per pupil spending throughout the school-age years could eliminate the attainment gaps between children from low-income and nonpoor families. This is a sizable effect. However, to put this effect size into perspective, the average family income was $31,925 for those from low-income families and $72,029 for those from nonpoor families, whereas in 2011 the 10th percentile of family income was $9,478 and the 90th percentile was $113,868 (all in 2000 dollars). The spending differences necessary to eliminate outcome difference between children from families at the 90th and the 10th percentiles of family income or between children from the poorest and the richest families are likely much larger than those we examine in our study. For all these reasons, the moderate convergence in outcomes across income groups observed over time in the aggregate are compatible with the magnitude of our estimated spending effects.

After Coleman et al. (1966), many have questioned whether money matters, and whether increased school spending can improve the lifetime outcomes of children from disadvantaged backgrounds. Our findings show that increased per pupil spending induced by state SFR policies did improve student outcomes and helped reduce the intergenerational transmission of poverty. Increased school funding alone may not guarantee improved outcomes, but our findings indicate that provision of adequate funding may be a necessary condition. Importantly, we find that how the money is spent may be important. As such, to be most effective it is likely that spending increases should be coupled with systems that help ensure spending is allocated toward the most productive inputs.

Northwestern University and NBER
UC Berkeley and NBER
Northwestern University

Supplementary Material

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

References

Angrist, Joshua D., Guido W. Imbens, and Alan Krueger, "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14 (1999), 57–67.
Angrist, Joshua D., and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106 (1991), 979–1014.
Ashenfelter, Orley, and Alan Krueger, "The American Economic Review," 84 (1994), 1157–1173.
Baicker, Katherine, and Nora Gordon, "The Effect of State Education Finance Reform on Total Local Resources," *Journal of Public Economics*, 90 (2006), 1519–1535.
Barrow, Lisa, Amy Claessens, and Diane Whitmore Schanzenbach, "The Impact of Chicago's Small High School Initiative," NBER Working Paper 18889, 2013.
Baum, Sandy, Jennifer Ma, and Kathleen Payea, "Education Pays 2013: The Benefits of Higher Education to Individuals and Society," Report, College Board, 2013.
Ben-Shalom, Yonatan, Robert Moffitt, and John Karl Scholz, "An Assessment of the Effectiveness of Anti-Poverty Programs in the United States," in *The Oxford Handbook of the Economics of Poverty*, Philip Jefferson, ed. (Oxford: Oxford University Press, 2012).
Betts, Julian R., "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth," *Review of Economics and Statistics*, 77 (1995), 231–250.
———., "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature," in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Gary Burtless, ed. (Washington, DC: Brookings Institution, 1996), 141–191.
Bloom, Howard S., and Rebecca Unterman, "Sustained Progress: New Findings about the Effectiveness and Operation of Small Public High Schools of Choice in New York City," MDRC Report, August 2013.
Card, David, and Alan B. Krueger, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100 (1992), 1–40.
Card, David, and Abigail A. Payne, "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores," *Journal of Public Economics*, 83 (2002), 49–82.
Carneiro, Pedro Manuel, and James J. Heckman, "Human Capital Policy," in *Inequality in America: What Role for Human Capital Policy?*, J. J. Heckman and A. Krueger, eds. (Cambridge, MA: MIT Press, 2003).
Carrell, Scott, and Susan Carrell, "Do Lower Student to Counselor Ratios Reduce School Disciplinary Problems?," *Contributions to Economic Analysis and Policy*, 5 (2006), 1463.
Chay, Kenneth Y., Jon Guryan, and Bhashkar Mazumder, "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth," NBER Working Paper 15078, 2009.
Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How Does Your

Kindergarten Classroom Affect Your Earnings? Evidence from Project Star," *Quarterly Journal of Economics*, 126 (2011), 1593–1660.
Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104 (2014), 2633–2679.
Coleman, J. S., E. Q. Campbell, C. J. Hobson, F. McPartland, A. M. Mood, F. D. Weinfeld, et al., *Equality of educational opportunity* Washington, DC: U.S. Government Printing Office, (1966).
Deming, David, "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start," *American Economic Journal: Applied Economics*, 1 (2009), 111–134.
Downes, Thomas A., and David Figlio, "School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?," Mimeo, University of Wisconsin, 1998.
Evans, William N., Craig Garthwaite, and Timothy J. Moore, "The White/Black Educational Gap, Stalled Progress, and the Long Term Consequences of the Emergence of Crack Cocaine Markets," NBER Working Paper 18437, 2012.
Fitzgerald, J., P. Gottschalk, and R. Moffitt, "An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33 (1998a), 251–299.
———, "The impact of attrition in the Panel Study of Income Dynamics on intergenerational analysis," *Journal of Human Resources*, 33 (1998b), 300–344.
Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek, "Long-Term Effects of Class Size," *Quarterly Journal of Economics*, 128 (2013), 249–285.
Fryer, Roland, Paul Heaton, Steven Levitt, and Kevin Murphy, "Measuring Crack Cocaine and Its Impact," *Economic Inquiry*, 51 (2013), 1651–1681.
Goodman, Joshua, "The Labor of Division: Returns to Compulsory Math Coursework," *Harvard University Mimeo*, 2012.
Grogger, Jeff, "Does School Quality Explain the Recent Black/White Wage Trend?," *Journal of Labor Economics*, 14 (1996), 231–253.
Guryan, Jonathan, "Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts," NBER Working Paper 8269, 2001.
Guryan, Jonathan, Erik Hurst, and Melissa Kearney, "Parental Education and Parental Time with Children," *Journal of Economic Perspectives*, 22 (2008), 23–46.
Haider, Steven, and Gary Solon, "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 96 (2006), 1308–1320.
Hanushek, Eric A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 49 (1986), 1141–1177.
———, "School Resources and Student Performance," in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Gary Burtless, ed. (Washington, DC: Brookings Institution, 1996).
———, "Spending on Schools," in *A Primer on American Education*, Terry Moe, ed. (Stanford, CA: Hoover Press, 2001).
———, "The Failure of Input Based Schooling Policies," *Economic Journal*, 113 (2003), 65–98.
Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103 (2014), 2052–2086.
Hightower, Amy M., Hajime Mitani, and Christopher B. Swanson, State Policies That Pay: A Survey of School Finance Policies and Outcomes (Bethesda, MD: Editorial Projects in Education and Pew Center on the States, 2010).
Howell, Penny L., and Barbara B. Miller, "Sources of Funding for Schools," *Future of Children*, 7 (1997), 39–50.
Hoxby, Caroline M., "Are Efficiency and Equity in School Finance Substitutes or Complements?," *Journal of Economic Perspectives*, 10 (1996), 51–72.
———, "All School Finance Equalizations Are Not Created Equal," *Quarterly Journal of Economics*, 116 (2001), 1189–1231.

Hoynes, Hilary W., Diane Whitmore Schanzenbach, and Douglas Almond, "Long Run Impacts of Childhood Access to the Safety Net," NBER Working Paper 18535, 2012.

Hyman, Joshua, "Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment," *University of Michigan mimeo*, 2014.

Jackson, C. Kirabo, "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence From the End of School Desegregation," *Journal of Labor Economics*, 27 (2009), 213–256.

———, "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina," NBER Working Paper 18624, 2012.

———, "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers," *Review of Economics and Statistics*, 95 (2013), 1096–1116.

Jackson, C. Kirabo, Rucker Johnson, and Claudia Persico, "The Effect of School Finance Reforms on the Distribution of Spending, Academic Achievement, and Adult Outcomes," NBER Working Paper 20118, 2014.

Jackson, M., and R. D. Mare, "Cross-Sectional and Longitudinal Measurements of Neighborhood Experience and Their Effects on Children," *Social Science Research*, 36 (2007), 590–610.

Jepsen, Christopher, Troske Kenneth, and Coomes Paul, "The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates," *Journal of Labor Economics*, 32 (2014) 95–121.

Johnson, Rucker C., "Race Differences in the Incidence and Duration of Exposure to Concentrated Poverty over the Life Course: Upward Mobility or Trapped in the Hood?," Goldman School of Public Policy Working Paper, U.C. Berkeley, 2008.

———, "Long-Run Impacts of School Desegregation and School Quality on Adult Attainments," NBER Working Paper 16664, 2011.

Katz, Lawrence, and David Autor, "Changes in the wage structure and earnings inequality," *Future of Children*, ch. 26 (1999), 1463–1555 in O. Ashenfelter and D. Card, eds. *Handbook of Labor Economics*, 3, Part A, Elsevier.

Kearney, Melissa S., Benjamin H. Harris, Elisa Jácome, and Lucie Parker, "Ten Economic Facts about Crime and Incarceration in the United States," Hamilton Project Policy Memo, 2014.

Krueger, Alan B., "Reassessing the View that American Schools Are Broken," New York Federal Reserve Economic Policy Brief, 1998.

Krueger, Alan B., and Diane Whitmore, "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project Star," *Economic Journal*, 111 (2001), 1–28.

Kunz, Jim, Marianne E. Page, and Gary Solon, "Are Single-Year Measures of Neighborhood Characteristics Useful Proxies for Children's Long-Run Neighborhood Environment?," *Economics Letters*, 79 (2003) 231–237.

Lindseth, Alfred, "Educational Adequacy Lawsuits: The Rest of the Story," paper presented at 50 Years after Brown: What Has Been Accomplished and What Remains to Be Done?, Cambridge, MA, April 2004.

Loeb, Susanna, and John Bound, "The Effect of Measured School Inputs on Academic Achievement: Evidence from the 1920s, 1930s and 1940s Birth Cohorts," *Review of Economics and Statistics*, 78 (1996), 653–664.

Ludwig, Jens, and Douglas L. Miller, "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics*, 122 (2007), 159–208.

Murnane, Richard J., "US High School Graduation Rates: Patterns and Explanations," *Journal of Economic Literature*, 51 (2013), 370–422.

Murray, Sheila E., William N. Evans, and Robert M. Schwab, "Education-Finance Reform and the Distribution of Education Resources," *American Economic Review*, 88 (1998), 798–812.

National Center for Education Statistics (NCES), "Digest of Education Statistics," U.S. Department of Education, 2012, available at http://nces.ed.gov/programs/digest/2012menu_tables.asp.

OECD, *Education at a Glance 2013: OECD Indicators* (Paris: OECD Publishing, 2013).

Owens, Ann, “Inequality in Children’s Contexts: The Economic Segregation of Households with and without Children,” Working Paper, Harvard University, 2015.

Papke, Leslie E., “The Effects of Spending on Test Pass Rates: Evidence from Michigan,” *Journal of Public Economics*, 89 (2005), 821–839.

Raphael, Steven, and Michael Stoll, eds., Do Prisons Make Us Safer? The Benefits and Costs of the Prison Boom, (New York: Russell Sage Foundation Press, 2009).

Reardon, Sean F., and Kendra Bischoff, “Income Inequality and Income Segregation,” *American Journal of Sociology*, 116 (2011), 1092–1153.

Reardon, S.F., “The Widening Income Achievement Gap,” *Educational Leadership*, 70 (2013), 10–16.

Reback, Randall, “Noninstructional Spending Improves Noncognitive Outcomes: Discontinuity Evidence from a Unique Elementary School Counselor Financing System,” *Education Finance and Policy*, 5 (2010), 105–137.

Roy, Joydeep, “Impact of School Finance Reform on Resource Equalization and Academic Performance: Evidence from Michigan,” *Education Finance and Policy*, 6 (2011), 137–167.

Schwartz, Amy Ellen, Leanna Stiefel, and Matthew Wiswall, “Do Small Schools Improve Performance in Large, Urban Districts? Causal Evidence form New York City,” *Journal of Urban Economics*, 77 (2013), 27–40.

Short, Kathleen, Timothy Smeeding, “Understanding Income-to-Threshold Ratios Using the Supplemental Poverty Measure,” U.S. Census Bureau Social, Economic, and Housing Statistics Division Working Paper 2012-18, 2012.

U.S. Department of Education, *Bureau of School Systems,* Public School Finance Programs of the United States and Canada: 1998–1999 (Washington, DC: U.S. Government Printing Office, 2001).

Waldfogel, Jane, Terry-Ann Craigie, and Jeanne Brooks-Gunn, “Fragile Families and Child Wellbeing,” *Future of Children,* 20 (2010), 87–112.

Watson, Tara, “Inequality and the Measurement of Residential Segregation by Income in American Neighborhoods,” *Review of Income and Wealth*, 3 (2009), 820–844.

ONLINE APPENDIX

THE EFFECTS OF SCHOOL SPENDING ON EDUCATIONAL AND ECONOMIC OUTCOMES: EVIDENCE FROM SCHOOL FINANCE REFORMS*

C. KIRABO JACKSON
NORTHWESTERN UNIV & NBER

RUCKER C. JOHNSON
UC-BERKELEY & NBER

CLAUDIA PERSICO
NORTHWESTERN UNIV

* Please direct correspondence to Kirabo Jackson (kirabo-jackson@northwestern.edu) and Rucker Johnson (ruckerj@berkeley.edu).

Appendix A

Table A1: Supreme Court Rulings on the Constitutionality of School Finance Systems from 1967-2010

State	First Case Name, Year	Second Case Name, Year	Third Case Name, Year	Fourth Case Name, Year	Fifth Case Name, Year
Alabama	Alabama Coalition for Equity v. Hunt; Harr v. Hunt, 1993				
Alaska	Kasayulie v. Alaska, 1999				
Arizona	Roosevelt v. Bishop, 1994	Roosevelt v. Bishop, 1997	Roosevelt v. Bishop, 1998	Flores v. Arizona, 2007	
Arkansas	Dupree v. Alma School District No. 30, 1983	Lake View v. Arkansas, 1994	Lake View School District, No. 25 v. Huckabee, 2002	Lake View School District, No. 25 v. Huckabee, 2005	
California	Serrano v. Priest, 1971	Serrano v. Priest, 1977	Eliezer Williams, et al., vs. State of California, et al, 2004		
Colorado	None				
Connecticut	Horton v. Meskill, 1978	Horton v. Meskill, 1982	Sheff v. O'Neill, 1995	Coalition for Justice in Education Funding, Inc v. Rell, 2010	
Delaware	None				
Florida	None				
Georgia	None				
Hawaii	None				
Idaho	Idaho Schools for Equal Educational Opportunity v. State, 1998	Idaho Schools for Equal Educational Opportunity v. State, 2005			
Illinois	None				
Indiana	None				
Iowa	None				
Kansas	Knowles v. State Board of Education, 1972	Montoy v. State, 2005			
Kentucky	Rose v. The Council for				

	Better Education, Inc., 1989				
Louisiana	None				
Maine	None				
Maryland	Bradford v. Maryland State Board of Education, 2005				
Massachusetts	Mc Duffy v. Secretary of the Executive Office of Education, 1993				
Michigan	Durant vs State of Michigan, 1997				
Minnesota	None				
Mississippi	None				
Missouri	Committee for Educational Equality v. Missouri, 1993				
Montana	Helena Elementary School District No. 1 v. State of Montana, 1989	Montana Rural Ed. Association v. Montana, 1993	Columbia Falls Public Schools v. State, 2005	Montana Quality Education Coalition v Montana, 2008	
Nebraska	None				
Nevada	None				
New Hampshire	Claremont New Hampshire v. Gregg, 1993	Claremont v. Governor, 1997	Claremont v. Governor, 1999	Claremont v. Governor, 2002	Londonderry School District v. New Hampshire, 2006
New Jersey	Robinson v. Cahill, 1973	Robinson v. Cahill, 1976	Abbott v. Burke, 1990	Abbott v. Burke, 1991	Abbott v. Burke, 1994
New Mexico	Zuni School District v. State, 1998				
New York	CFE v. State, 2003	CFE v. State, 2006			
North Carolina	Leandro v. State, 1997	Leandro v. State, 2004			
North Dakota	None				
Ohio	DeRolph v. Ohio, 1997	DeRolph v. Ohio, 2000	DeRolph v. Ohio, 2002		
Oklahoma	None				
Oregon	Pendleton School District				

	v. State of Oregon, 2009				
Pennsylvania	None				
Rhode Island	None				
South Carolina	Abbeville County School District v. State, 2005				
South Dakota	None				
Tennessee	Tennessee Small School Systems v. McWheter, 1993	Tennessee Small School Systems v. McWheter, 1995	Tennessee Small School Systems v. McWheter, 2002		
Texas	Edgewood Independent School District v. Kirby, 1989	Edgewood Independent School District v. Kirby, 1991	Carrollton-Farmers v. Edgewood, 1992	West Orange-Cove Consolidated ISD v. Nelson, 2004	
Utah	None				
Vermont	Brigham v. State, 1997				
Virginia	None				
Washington	Seattle School District No. 1 of King County v. State, 1977	Seattle II, 1991	Federal Way School District v. State of Washington, 2007		
West Virginia	Pauley v. Kelly, 1979	Pauley v. Bailey, 1984	Pauley v. Gainer, 1995		
Wisconsin	Buse v. Smith, 1976				
Wyoming	Washakie v. Herschler, 1980	Campbell v. State, 1995	Campbell II, 2001		

Appendix B: ***Coverage of School Districts in our Data***

Previous historical data on per-pupil expenditures was only available in a readily usable format via the *Census of Governments: School System Finance (F-33) File* (U.S. Bureau of the Census, Department of Commerce). The Census of Governments previously was only conducted in years that end in a two or seven, so at the time when many important papers on SFRs were written, there were many years of missing data. In addition, until recently the earliest available F-33 data was for the year 1972. As a result, it was previously impossible to model per-pupil spending and spending inequality annually over time, so many authors (e.g., MES, Card and Payne), operating under the Common Trends Assumption, assumed that trends in per-pupil spending were linear. Due to these limitations, previous papers on school finance reforms were also unable to look at how the exact timing of reforms affected per-pupil expenditure and spending inequality within a state.

Our data from the Historical Database on Individual Government Finances (INDFIN) represents the Census Bureau's first effort to provide a time series of historically consistent data on the finances of individual governments. This database combines data from the *Census of Governments Survey of Government Finances (F-33),* the National Archives, and the *Individual Government Finances Survey*. The School District Finance Data FY 1967-91 is available annually from 1967 through 1991. It contains over one million individual local government records, including counties, cities, townships, special districts, and independent school districts. The INDFIN database frees the researcher from the arduous task of reconciling the many technical, classification, and other data-related changes that have occurred over the last 30 years. For example, this database includes corrected statistical weights that have been standardized across years, which had not been done previously. Furthermore, although most governments retain the ID number they are assigned originally, there are circumstances that result in a government's ID being changed. Since a major purpose of the INDFIN database is tracking government finances over time, it is critical that a government possess the same ID for all years (unless the ID change had a major structural cause). For example, All Alaska IDs were changed in the 1982 Census of Governments. In addition, new county incorporations, where governments in the new county area are re-assigned an ID based on the new county code (e.g., La Paz County, AZ), cause ID changes. Thus, if a government ID number was changed, the ID used in the database is its current GID number, including those preceding the cause of the change, so that the ID is standardized across years.

In addition to standardizing the data, the Census Bureau has corrected a number of errors in the INDFIN database that were previously in other sources of data. For example, for fiscal years 1974, 1975, 1976 and 1978 the school district enrollment data that had previously been released were useless (either missing or in error for many records). Thus, in August 2000, these missing enrollment data were replaced with those from the employment survey individual unit files. This enables us to more accurately compute per-pupil expenditures for those years. In addition, source files before fiscal 1977 were in whole dollars rather than thousands. This set a limit on the largest value any field could hold. If a figure exceeded that amount, then the field contained a special "overflow" flag (999999999). Few governments exceeded the limit (Port Authority of NY and NJ and Los Angeles County, CA are two that did). For the INDFIN database, actual data were substituted for the overflow flag. Finally, in some cases the Census revised the original data in source files for the INDFIN database. In some cases, official revisions were never applied to the data files. Others resulted from the different environment and operating practices under which source files were created. Finally, some extreme outliers were identified and corrected (e.g., a keying error for a small government that ballooned its data).

The Common Core of Data (CCD) School District Finance Survey (F-33) consists of data submitted annually to the National Center for Education Statistics (NCES) by state education agencies (SEAs) in the 50 states and the District of Columbia. The purpose of the survey is to provide finance data for all local education agencies (LEAs) that provide free public elementary and secondary education in the United States. Both NCES and the Governments Division of the U.S. Census Bureau collect public school system finance data, and they collaborate in their efforts to gather these data. The Census of Governments, which was recorded every five years until 1992, records administrative data on school spending for every district in the United States. After 1992, the Public Elementary-Secondary Education Finances data were recorded annually with data available until 2010. We combine these data sources to construct a long panel of annual per-pupil spending for each school district in the United States between 1967 and 2010.

Per-pupil spending data from before 1992 is missing for Alaska, Hawaii, Maryland, North Carolina, Virginia, and Washington, D.C. Per-pupil spending data from 1968 and 1969 is missing for all states. Spending data in Florida was also missing for 1975, 1983, 1985-1987, and 1991. Spending data in Kansas was also missing for 1977 and 1986. Spending data in Mississippi was also missing for 1985 and 1988. Spending data in Wyoming was also missing for 1979 and 1984. Spending data for Montana is missing in 1976, data for Nebraska is missing in 1977, and data for Texas is missing in 1991. Where there was only a year or two of missing per-pupil expenditure data, we filled in this data using linear interpolation.

Figure A1 below shows the number of district observations in our data for each year. The bars highlighted in red are the census of government years employed in previous national studies of school finance reforms (e.g. Card and Payne 2002, Hoxby 2001, Murray Evans and Schwab 1998). While the coverage of the data we use is arguably better that that used previously, it is not perfect. As shown in Appendix figure A1, for years, 1967, 1970, 1971, 1972, 1972, 1975, 1976, and 1978 only about 40 percent of districts are present (often larger districts). After 1979 almost all districts are included. To assuage concern that our results are affected by the composition of districts for these few years with incomplete coverage, we estimate the main event-study models using only data after 1979. As one can see in figures B2 and B3, the pattern of results is similar to that for the full sample (B2) when the sample is restricted to only observations after 1979 (B3).

Figure B1: *The number of district observations for each year.*

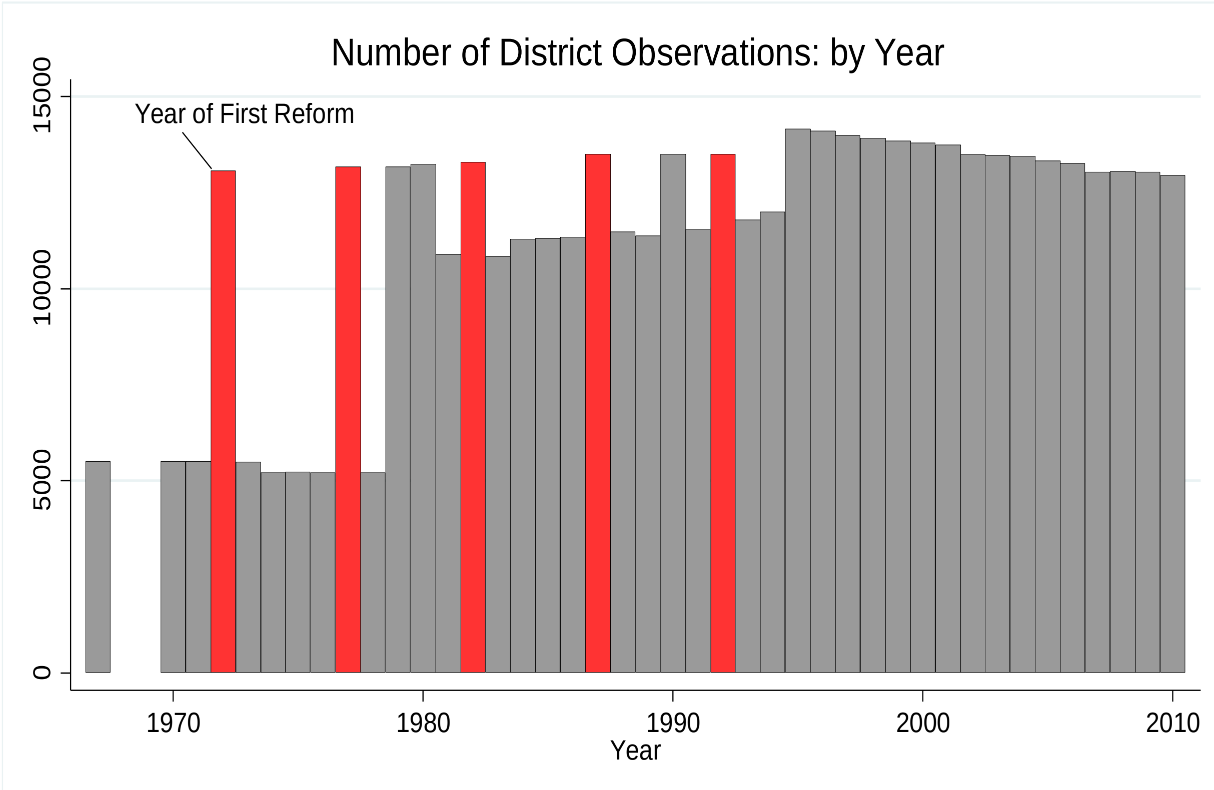


Figure B2: *The Effect of a Court-ordered Reform on Per-Pupil Spending by Pre-Reform Spending Quartile* (All Years)

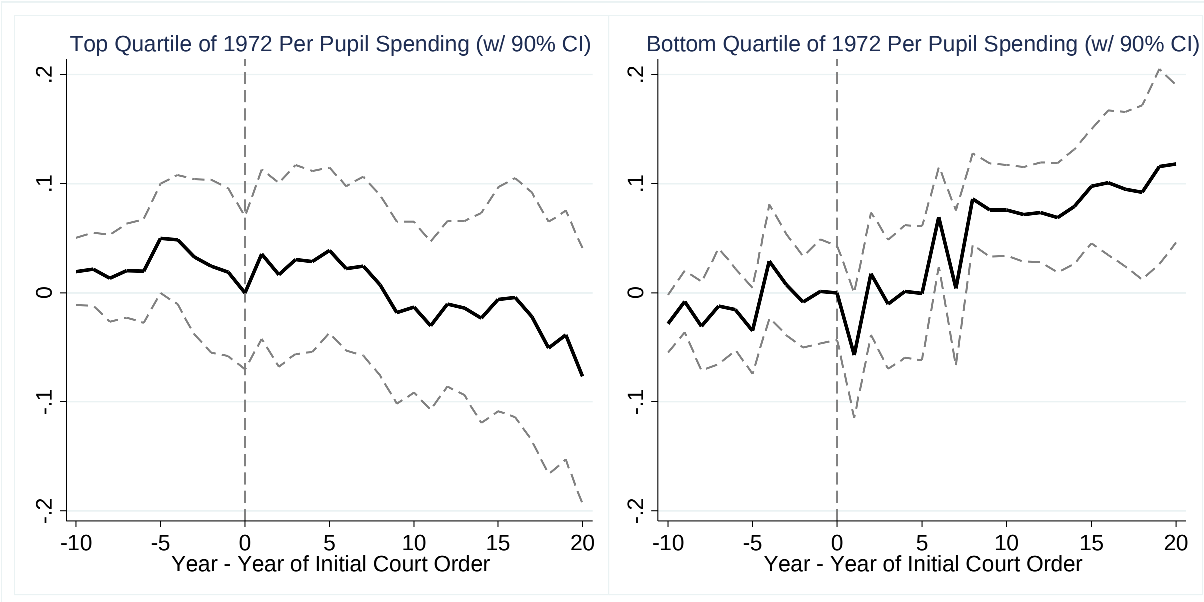
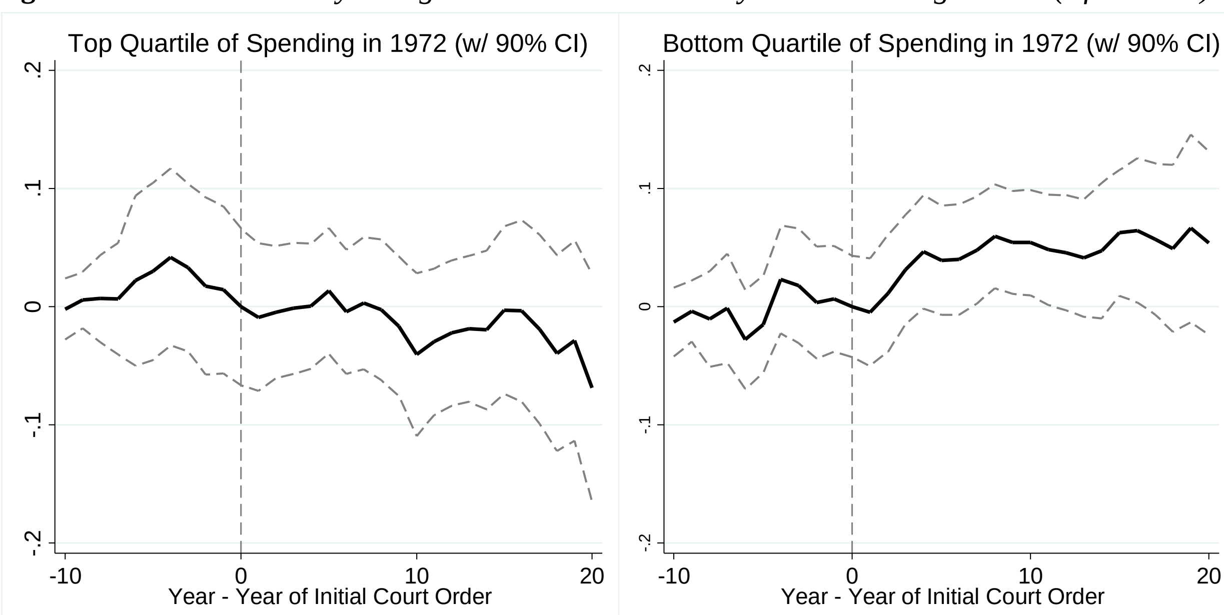


Figure B3: *The Event-study Using All the Data versus only Good Coverage Years (After 1979)*



Appendix C: *Data on School Finance Reforms*

Due to great interest on the topic, the timing of school finance reforms (SFRs) has been collected in various places. Data on the exact timing and type of court-ordered and legislative SFRs was obtained from Public School Finance Programs of the Unites States and Canada (PSFP), National Access Network's state by state school finance litigation map (2011), from Murray, Evans, and Schwab (1998), Hoxby (2001), Card and Payne (2002), Hightower et al (2010), and Baicker and Gordon (2004). The most accurate information on school finance laws can be derived from the PSFP, which provides basic information and references to the legislation and court cases challenging them (Hoxby 2001). In most cases, data from these sources are consistent with each other. Where there are discrepancies we often defer to PSFP, but also consulted LexisNexis and state court and legislation records.

There were discrepancies in reported timing of overturned court cases in several states: Connecticut (Hoxby states the decision was made in 1978, but Card and Payne report it was made in 1977), Kansas (Hoxby states 1976, but PSFP and ACCESS report 1972), New Jersey (Card and Payne state 1989, but PSFP says 1990), Washington (Murray, Evans, and Schwab, Hoxby, and Card and Payne report 1978, but PSFP reports 1977), Wyoming (Hoxby says 1983, but Card and Payne and Murray, Evans, and Schwab report 1980). We researched each case by name to discover the true date of the decision.

Using a policy survey conducted during the 2008-2009 school year, a recent study by Hightower et al (2010) provides a description of state finance policies and practices. This study was used to verify whether there had been any changes to state funding formulas between 1998 and 2009. We only collected information on the first five court cases per state in which the state found the school funding system unconstitutional. There were only three states with five or more court cases overruling the funding system (New Hampshire, New Jersey, and Texas). In addition, we only collected information on the first four court cases per state in which states upheld the school funding system. There were only four states with four or more court cases in which the school funding system was upheld (Illinois, New York, Oregon, and Pennsylvania).

Information on whether or not a state funding formula had a MFP, flat grant formula, variable matching grant scheme, recapture provision, spending limit, power equalization scheme, local-effort equalization scheme, or full state funding came from *PSFP* (1998) and was verified using Card and Payne (2002) and Hightower et al (2010). We defined MFPs, flat grant formulas, and variable matching grant schemes in the same way as Card and Payne did in their 2002 study. We defined power equalization, local-effort equalization, and full state funding in the same way as the EPE study (Hightower, Mitani and Swanson 2010). Each element of a state funding formula was coded as a dichotomous variable. For example, MFP is a dichotomous variable that is equal to one in the year and all subsequent years in which a state's finance system had a MFP plan in place. MFP was set equal to zero in all years prior to the state's funding system having a MFP in place, or if a state never implemented a MFP. Information on the timing of spending and tax limits came from Downes and Figlio (1998). We also supplemented this with data from *PSFP* for years after those covered in Downes and Figlio (1998).

Appendix D: *A More Detailed Discussion about the Reform Types Used*

As pointed out in Hoxby (2001), the effect of a reform on school spending depends on the type of school funding formula introduced by the reform. We follow the typology outlined by Jackson, Johnson, and Persico (2014b) and categorize funding formulas into four main types; foundation, equalization, reward for local effort, and spending limits. To show how these formula types affect school spending, we estimate the same event-study model as described in [1] where timing of the specific formula change is used (e.g. year relative to the first introduction of a foundation plan). Even though we discus each reform type in turn, the event-study figures for each reform type are based on a model that includes the effect of all four reform types simultaneously.

First we consider formulas that impose **spending limits**. Under such plans, the state imposes a limit on how much a district may spend on education. In addition, some plans take away all tax revenues raised above a certain amount. A key feature of such formulas is that at the limit districts face a zero inverted tax price – that is, a district receives zero additional dollars from raising one dollar in local revenue. Figures D2 and D3 show the event-study for state plans that impose spending limits for districts in the top and bottom of the median income distribution in 1963. Spending limits reduce per-pupil spending for all districts in the long run with the most pronounced effect in the high spending and more affluent districts.

On the other side of the policy spectrum are "**reward for local effort**" formulas that encourage local districts to increase per-pupil spending with matching funds. These formulas affect tax prices directly. These plans provide greater incentives for lower-income and low-wealth districts to increase taxes by allowing such districts to have more than one dollar in spending for each dollar raised in taxes.[1] Figures D2 and D3 provide the event-study for this kind of formula change. There is evidence of a downward pre-trend. However, relative to trend, there is a clear spending increase for all districts (by income or spending levels prior to reforms).

The third kind of formula we consider are **foundation formulas**. These plans establish a foundation level of per-pupil spending, estimate a district's required local contribution to fund this foundation level based on income and wealth levels in the district, and provide the difference between the expected contribution and the foundation level. These plans do not affect tax prices, but do redistribute resources to provide extra funding to low-income/low-wealth districts. Figures D2 and D3 present the event-study for foundation plans. Foundation plans tended to be introduced in states that saw increased school spending. While the effect on spending levels is difficult to discern given the pre-existing trends, the gap in spending associated with these reforms between the low- and high-income districts (and also high and low spending district) was reduced.

Finally, we turn to **equalization formulas**. Equalization plans redistribute locally raised and state funds. They provide extra funding to low- income or wealth districts while possibly taking money away from high-income or wealthy districts. These plans do not affect tax prices directly although they may provide incentives to alter the tax base. Figures D2 and D3 present the event-study for equalization plans. There is evidence of a positive pre-trend for both district types. However, the gaps in spending between low- and high-income districts was stable prior to reforms. While equalization plans as small effect on the spending gap by median income, spending gaps by prior spending was reduced by these plans.

In forming our instrument we also use the passage of "**adequacy based court-ordered reforms"**. There were two waves of court-ordered reforms (see Figure D4). In the first wave,

---

[1] For example, in Georgia, school districts at or below 75 percent of the state average property tax wealth level receive equalization funding in proportion to the number of mills they raise above the required five mill.

known as "equity cases," proponents of state funding argued that local financing violated the responsibility of the state to provide a quality education to all children. They asserted that public education was a "fundamental interest" for equal protection purposes and thus could not be distributed unequally within a state based on geography absent any "compelling state interest." The motivation was that "poor" school districts had little property wealth to tax in order to support their local schools, while "rich" school districts had much more at their disposal. As such, despite the greater tax effort by residents in these poor school districts, they would end up with less money per-pupil because of the difference in assessed wealth. Cases during the second wave of successful challenges were argued on adequacy grounds. "**Adequacy cases**" rely on the fact that virtually all states have a constitutional provision requiring the state to provide some level of free education for children (Lindseth, 2004). These cases were argued on the ground that prevailing low levels of educational resources in certain districts (typically low-income areas) violated the state's duty to provide the necessary educational opportunities guaranteed by the state constitution.

Figure D1: *Event-study Estimates of The Effect of Initial Court-ordered Reform and Adequacy Based Court-ordered Reforms on Per-pupil Spending: By Median Income quartile in 1969*

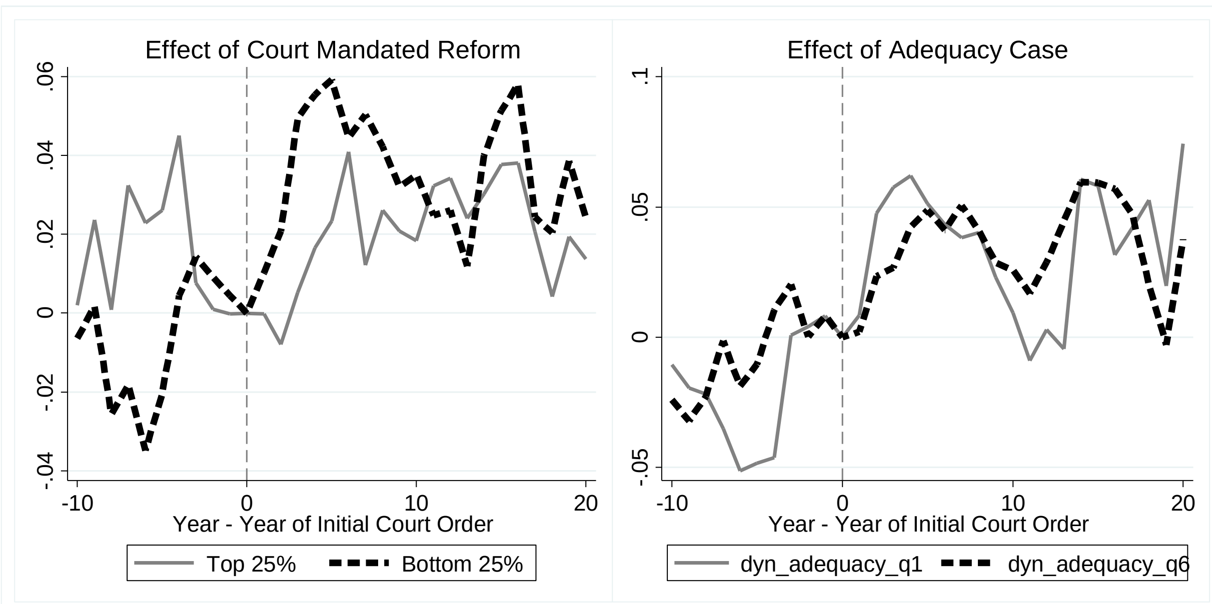
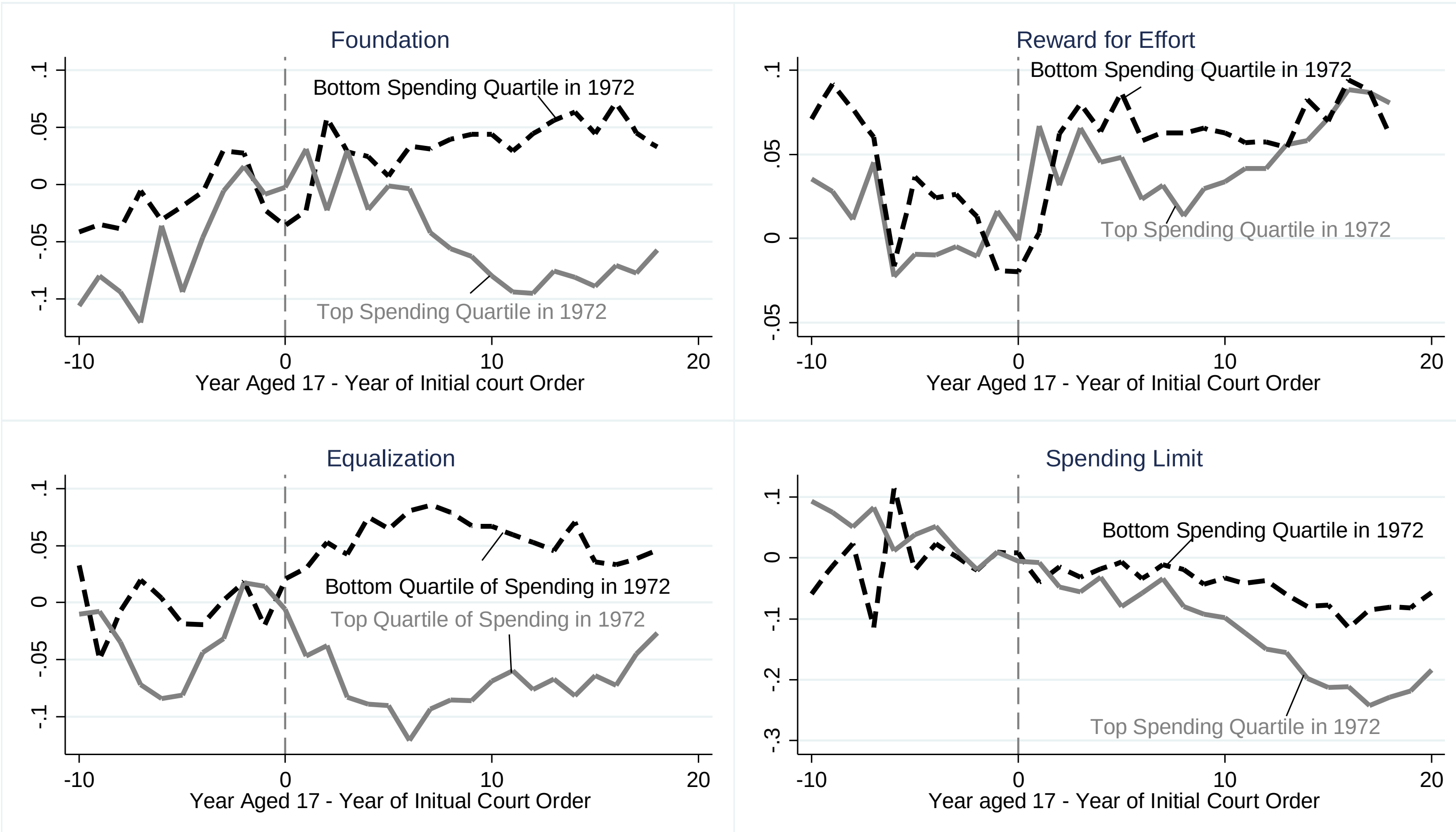
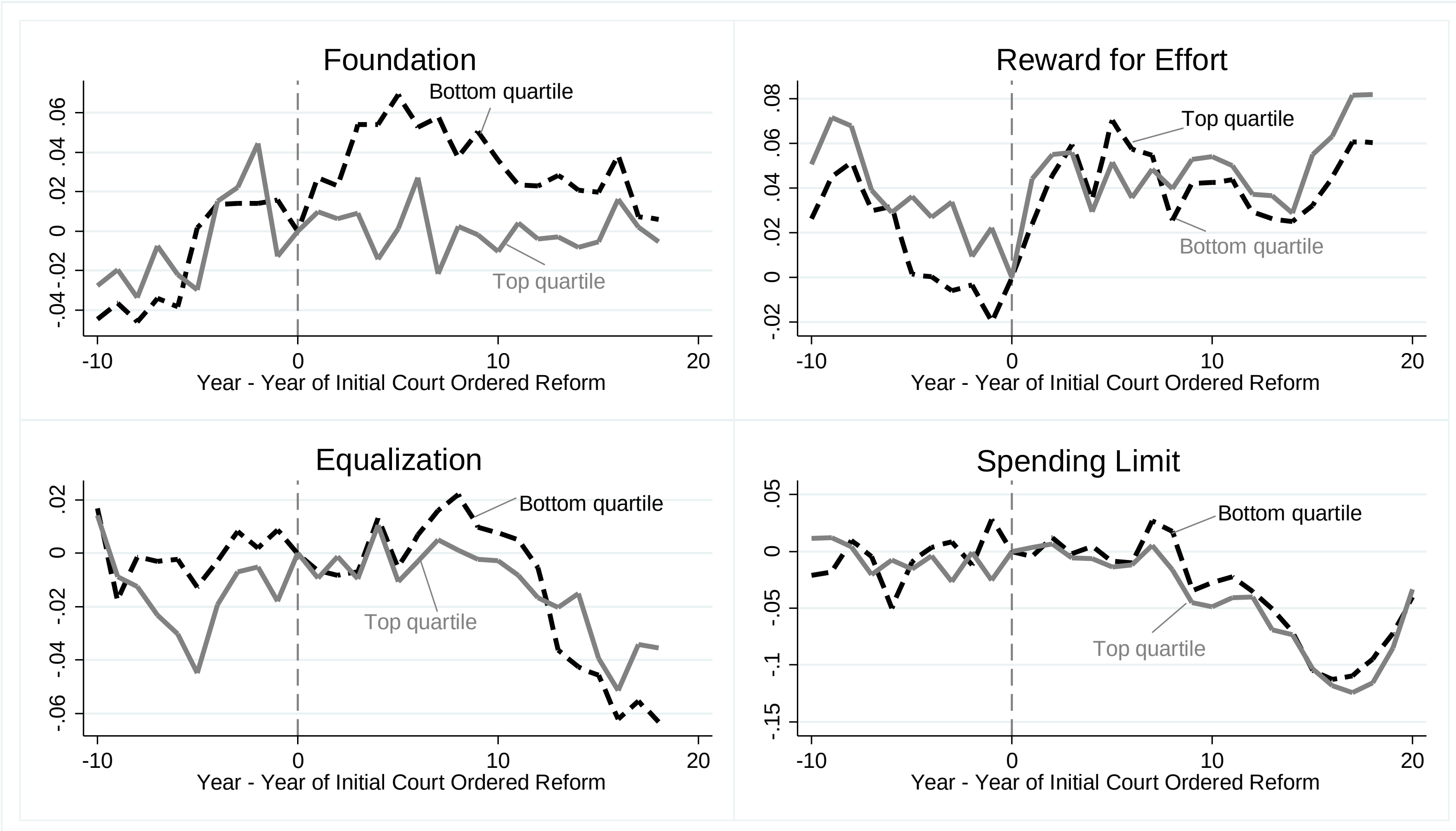


Figure D2: *Event-study Estimates of The Effect of Adopting Various Formula Types on Per-pupil Spending: By Spending quartile in 1972*



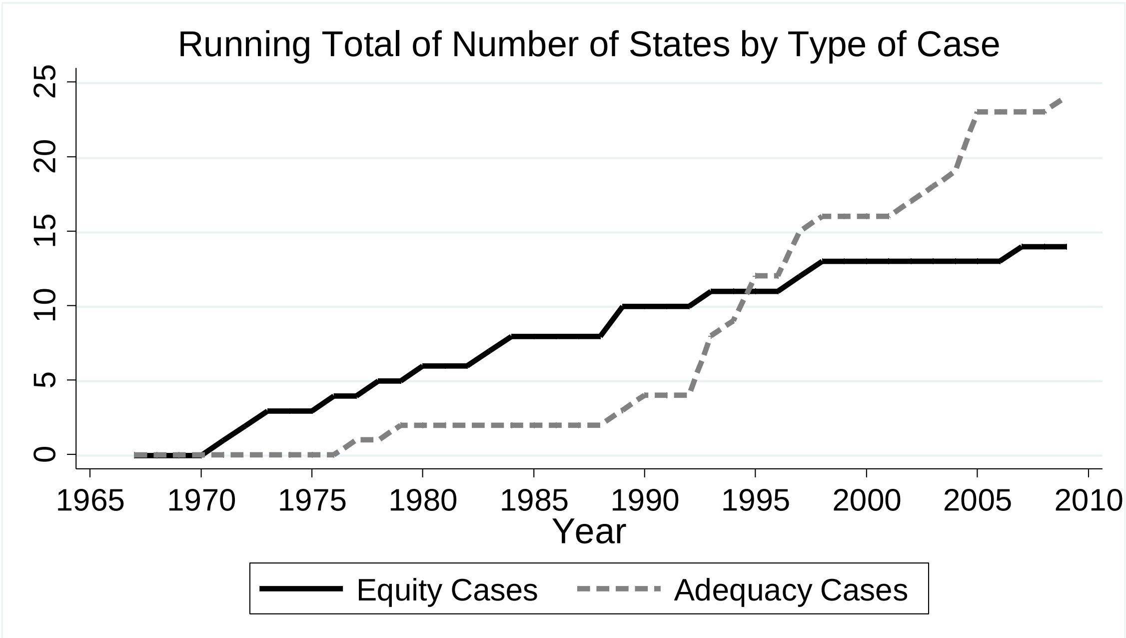
Data: The sample includes all school districts in the United States between the years of 1967 and 2010. The sample is made up of 483,047 district-year observations. Each district is weighted by average enrollment for the full sample.
Model: These plots present the estimated event time coefficients of a regression on per-pupil spending at the district level on year fixed effects, district fixed effects, and the percentile group of the district in the state distribution of median income interacted with a full set of event-time indicator variables from 10 years prior to 19 years after the first court-mandated reform. The event-study plots are shown for the top and bottom 25 percent of districts in the state distribution of per-pupil spending in 1972. The event time plot has been re-centered at zero for the 10 pre-reform years so that the estimated coefficients represent the change in spending relative to the levels that persisted in the 10 years prior to the first reform.

Figure D3: *Event-study Estimates of The Effect of Adopting Various Formula Types on Per-pupil Spending: By Income quartile in 1969*



<u>Data:</u> The sample includes all school districts in the United States between the years of 1967 and 2010. The sample is made up of 483,047 district-year observations. Each district is weighted by average enrollment for the full sample.
<u>Model:</u> These plots present the estimated event time coefficients of a regression on per-pupil spending at the district level on year fixed effects, district fixed effects, and the percentile group of the district in the state distribution of median income interacted with a full set of event-time indicator variables from 10 years prior to 19 years after the first court-mandated reform. The event-study plots are shown for the top and bottom 25 percent of districts in the state distribution of median family income in 1969. The event time plot has been re-centered at zero for the 10 pre-reform years so that the estimated coefficients represent the change in spending relative to the levels that persisted in the 10 years prior to the first reform.

Figure D4: *Type of Cases over Time*



Appendix E: *Matching PSID Individuals to their Childhood School Districts*

In order to limit the possibility that school district boundaries were drawn in response to pressure for SFRs, we utilize 1969 school district geographies. The "69-70 School District Geographic Reference File" (Bureau of Census, 1970) relates census tract and school district geographies. For each census tract in the country, it provides the fraction of the population that is in each school district. Using this information, we aggregate census tracts to 1970 district geographies with Geographic Information Systems (GIS) software. We assign census tracts from 1960, 1980 and 1990 to school districts using this resulting digital map based on their centroid locations.

To construct demographic information on 1969-1970-definition school districts, we compile census data from the tract, place, school district and county levels of aggregation for 1960, 1970, 1980 and 1990. We construct digital (GIS) maps of 1970 geography school districts using the 1969-1970 School District Geographic Reference File from the Census. This file indicates the fraction by population of each census tract that fell in each school district in the country. Those tracts split across school districts we allocated to the school district comprising the largest fraction of the tract's population. Using the resulting 1970 central school district digital maps, we allocate tracts in 1960, 1980 and 1990 to central school districts or suburbs based on the locations of their centroids. The 1970 definition central districts located in regions not tracted in 1970 all coincide with county geography which we use instead.

Appendix F: 2SLS Effects Using the Absolute Level of Per-pupil Spending (as opposed to logs)

Table F1: *2SLS Effects of Spending Levels on Outcomes*

	Years of Education		Prob(High School Grad)		Ln(Wage), age 20-45		Ln(annual Family Income), age 20-45		Prob(Adult Poverty), age 20-45	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Spending$_{(\text{age 5-17})}$	0.5859***		0.1338***		0.1314***		0.1657***		-0.0499***	
	(0.1607)		(0.0276)		(0.0361)		(0.0528)		(0.0132)	
Spending$_{(\text{age 5-17})}$*Low Income		0.7398***		0.1893***		0.1756***		0.3033***		-0.1170***
		(0.2225)		(0.0517)		(0.0552)		(0.0669)		(0.0228)
Spending$_{(\text{age 5-17})}$*Non Poor		0.1034		0.0459		0.0888		0.0300		0.0052
		(0.2492)		(0.0340)		(0.0814)		(0.0761)		(0.0155)
Number of person-year observations	--	--	--	--	106,545	106,545	151,349	151,349	151,756	151,756
Number of Individuals	15,353	15,353	15,353	15,353	13,183	13,183	14,730	14,730	14,737	14,737
Number of Childhood Families	4,586	4,586	4,586	4,586	4,454	4,454	4,588	4,588	4,588	4,588
Number of School Districts	1,409	1,409	1,409	1,409	1,395	1,395	1,414	1,414	1,414	1,414

Robust standard errors in parentheses (clustered at school district level) *** p<0.01, ** p<0.05, * p<0.10

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative.

Models: The school spending variable is in $1000s (CPI-deflated real 2000 dollars), so that a one-unit change represents a $1,000 increase experienced in each school-age year between ages 5-17. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of Spend$_d$).

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; In models by childhood poverty status all variables are interacted with childhood poverty status.

Appendix G: 2SLS Effects with Standard Errors Clustered at the Childhood State Level

Table G1: *2SLS Effects of Spending Levels on Outcomes*

	Years of Education		Prob(High School Grad)		Ln(Wage), age 20-45		Ln(annual Family Income), age 20-45		Prob(Adult Poverty), age 20-45	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$Ln(PPE_d)_{(age\ 5\text{-}17)}$	3.1488***		0.7053***		0.7743***		0.9819***		-0.2678***	
	(0.5224)		(0.1151)		(0.1427)		(0.2362)		(0.0495)	
$Ln(PPE_d)_{(age\ 5\text{-}17)}$ × Low		4.5899***		0.9878***		0.9598***		1.7146***		-0.6132***
		(0.9280)		(0.2075)		(0.1975)		(0.3343)		(0.1325)
$Ln(PPE_d)_{(age\ 5\text{-}17)}$ × Non Poor		0.7156		0.2470**		0.5525+		0.2021		0.0385
		(1.1309)		(0.1204)		(0.3560)		(0.3429)		(0.1269)
Number of person-year obs	--	--	--	--	106,545	106,545	151,349	151,349	151,756	151,756
Number of Individuals	15,353	15,353	15,353	15,353	13,183	13,183	14,730	14,730	14,737	14,737
Number of Childhood Families	4,586	4,586	4,586	4,586	4,454	4,454	4,588	4,588	4,588	4,588
Number of School Districts	1,409	1,409	1,409	1,409	1,395	1,395	1,414	1,414	1,414	1,414

Robust standard errors in parentheses (clustered at the childhood state level) *** p<0.01, ** p<0.05, * p<0.10

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative.

Models: The key treatment variable, $Ln(PPE_d)_{(age\ 5\text{-}17)}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below.

2SLS 1: The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the state distribution of per pupil school spending in 1972).

2SLS 2: The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$).

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; In models by childhood poverty status all variables are interacted with childhood poverty status.

Appendix H: *Cross Tabulations of Quantiles of Median Family Income in 1969, Per-pupil spending in 1972 and Spend$_d$*

Note: All numbers presented are row percentages

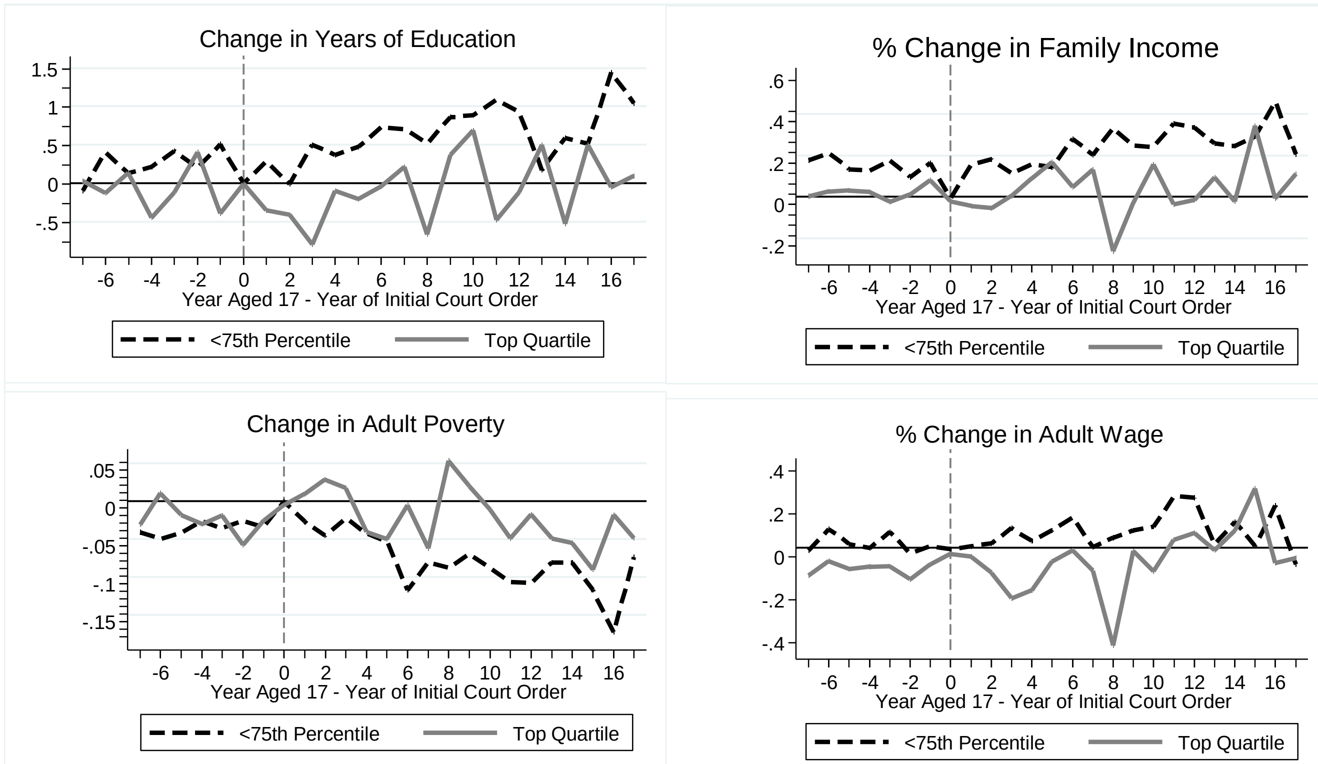
		Quantile of Spending in 1972			
		1	2	3	4
Quantile of Spend$_d$	1	15.81	15.1	18.78	**50.31**
	2	11.33	16.6	**32.92**	39.14
	3	28.86	**29.72**	24.34	17.08
	4	**39.41**	30.28	21.65	8.66

		Quantile of Income in 1969			
		1	2	3	4
Quantile of Spend$_d$	1	2.41	10.95	27.24	**59.4**
	2	15.1	9.79	**40.32**	34.79
	3	22.93	17.34	27.87	**31.86**
	4	**48.97**	11.93	22.17	16.92

		Quantile of Income in 1969			
		1	2	3	4
Quantile of Spending in 1969	1	**40.69**	11.89	22.81	24.61
	2	26.66	14.13	25.43	**33.78**
	3	22.27	12.89	27.29	**37.54**
	4	14.13	11.58	32.54	**41.76**

Appendix I

Figure I1: *Event-study Effects on Initial Court-ordered Reform by Spending Quartile in 1972*



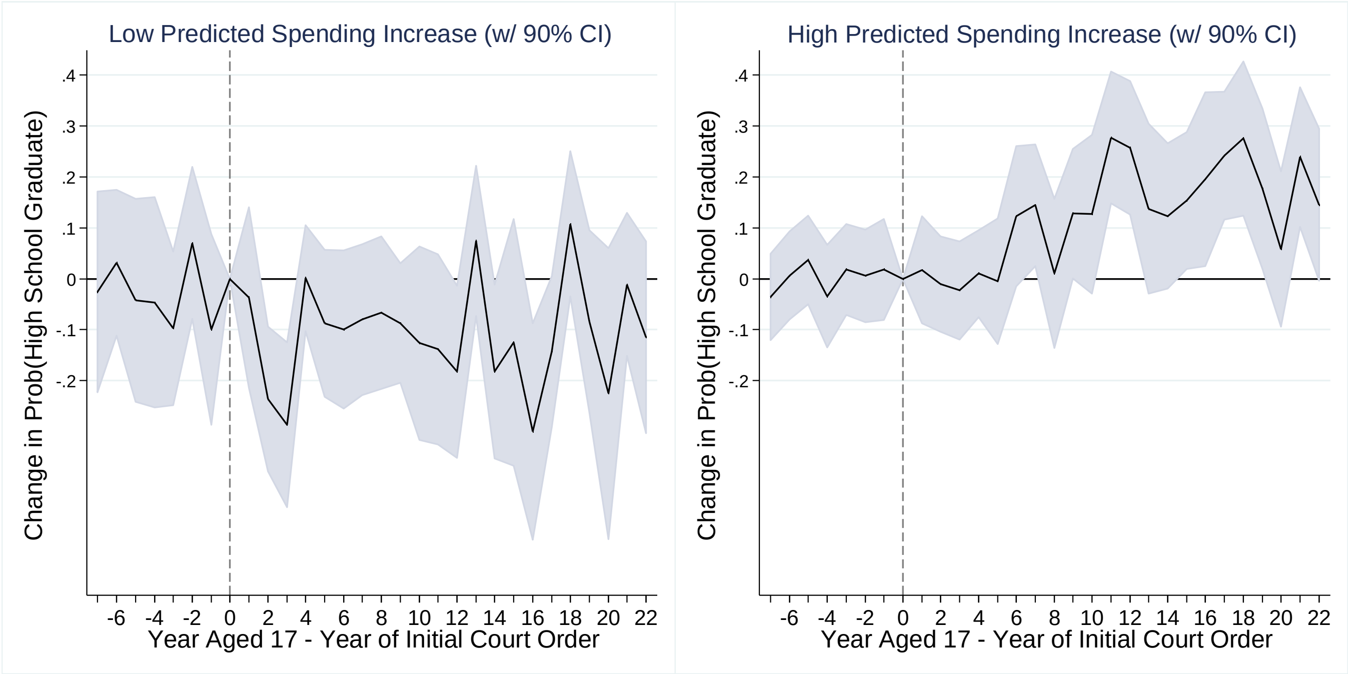
Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states). Sampling weights are used so that the results are nationally representative.

Models: The event-study plot is based on indicator variables for the number of school-age years of exposure to a court ordered SFR interacted with an indicator for whether the district was in the top quartile of the state distribution of per-pupil spending in 1972. Results are based on non-parametric event-study models that include school district fixed effects, race×census division×birth cohort fixed effects and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race) each interacted with linear cohort trends.

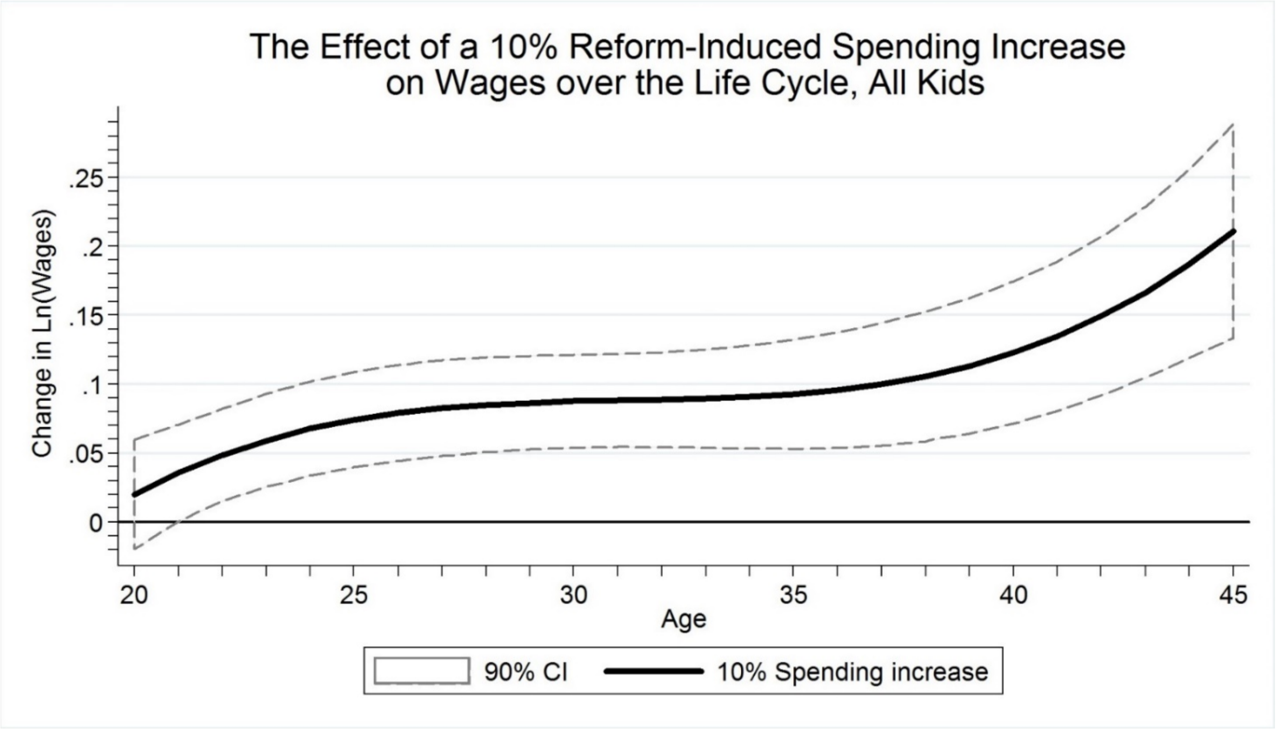
Appendix J: *Additional Analysis from Section V*

Appendix Figure J1: *Effect of Court-Ordered School Finance Reform on Likelihood of High School Graduation*



Note: High predicted spending increase refers to districts in reform states with $Spend_{d}$.>0 and low predicted spending increase refers to districts in reform states with $Spend_{d}$.≤0. Roughly two-thirds of districts in reform states had $Spend_{d}$.>0.
Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states). Sampling weights are used so that the results are nationally representative.
Models: The event-study plot is based on indicator variables for the number of school-age years of exposure to a court ordered SFR interacted whether the district is predicted to experience a spending increase due to reforms ($Spend_{d}$.>0) or not. Results are based on non-parametric event-study models that include school district fixed effects, race×census division×birth cohort fixed effects and additional controls.
Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race) each interacted with linear cohort trends.

Appendix Figure J2: *Effect of a Ten Percent Reform Induced Spending Increase on Log Wages: By age*



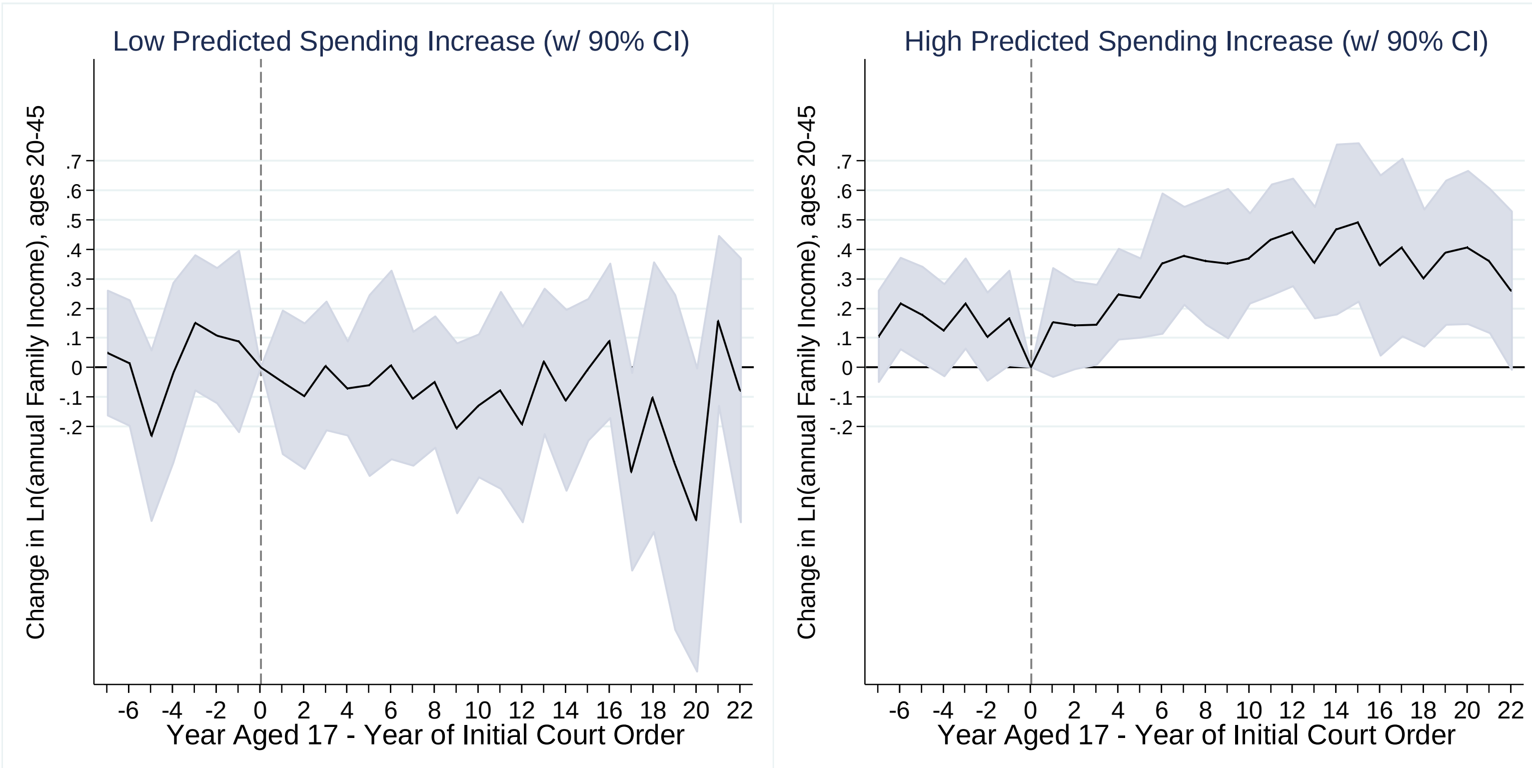
Notes: This figure plots the marginal effect of a 10 percent spending increase based on the interaction of per-pupil district spending and a cubic in age from the preferred 2SLS model. One rejects a linear model at the 1 percent level.

Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states). Sampling weights are used so that the results are nationally representative.

Models: The key treatment variables are the natural log of school-age per pupil spending interacted with age, age squared and age cubed. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$), all interacted with age, age squared and age cubed.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race) each interacted with linear cohort trends; In models by childhood poverty status all variables are interacted with childhood poverty status.

Appendix Figure J3: *Effect of Court-Ordered School Finance Reform on ln(Annual Family Income)*



Note: High predicted spending increase refers to districts in reform states with $Spend_{d}$.>0 and low predicted spending increase refers to districts in reform states with $Spend_{d}$.≤0. Roughly two-thirds of districts in reform states had $Spend_{d}$.>0.

Data: PSID geocode Data (1968-2011), matched with childhood school and district characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011, (N=15,353 individuals from 1,409 school districts (1,031 child counties, 50 states). Sampling weights are used so that the results are nationally representative.

Models: The event-study plot is based on indicator variables for the number of school-age years of exposure to a court ordered SFR interacted whether the district is predicted to experience a spending increase due to reforms ($Spend_{d}$.>0) or not. Results are based on non-parametric event-study models that include school district fixed effects, race×census division×birth cohort fixed effects and additional controls.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race) each interacted with linear cohort trends.

Appendix Table J1: *2SLS Models of the Effect on Wages using only a Single Year of Earnings Data*

	Point in Time Wage Effects Versus Wages over Multiple Time Periods		
	Using 2001 survey Data Only	Using 2010 survey Data Only	All Years
	1	2	3
$Ln(PPE_d)_{(age\ 5-}$	0.3729	0.4852	0.7743***
	(0.5405)	(0.5467)	(0.1959)

Robust standard errors in parentheses (clustered at school district level) *** $p<0.01$, ** $p<0.05$, * $p<0.10$

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative.

Models: The key treatment variable, $Ln(PPE_d)_{(age\ 5\text{-}17)}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the state distribution of per pupil school spending in 1972).

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; In models by childhood poverty status all variables are interacted with childhood poverty status.

Appendix K: *Falsification Placebo Test*

We probed the robustness of these 2SLS estimates further in several ways. First, as a placebo falsification test using the 2SLS models, we estimate the marginal effect of school spending during non-school-age years. That is, we estimate 2SLS models similar to equations [1] and [2] where, in addition to including instrumented school spending between the ages 5 and 17, we also include instrumented school spending between the ages of 20 and 24 (when there should be no effect) in the same model. To isolate exogenous changes in school spending for the different age ranges we use an instrument for exposure during the respective age ranges. As before, we instrument for school spending between ages 5 and 17 (school-age years) with the number of years of exposure between ages 5 and 17 interacted with the quartile of the district in the distribution of $Spend_d$. Similarly, we instrument for school spending between ages 20 and 24 (post-school-age years) with the number of years of exposure between ages 20 and 24 interacted with the quartile of the district in the distribution of $Spend_d$. The first-stage yields F-statistics above 10 for both of the endogenous regressors. If the effects documented are truly reflective of the causal effects of school spending, significant effects should be present during school-age years with no corresponding significant effects for non-school-age years.[2] While the placebo estimates are noisy, the results of the placebo tests presented in Table K1 support a causal interpretation. For all outcomes, there are statistically significant effects of reform-induced spending during school-age years and no statistically significant effect of school spending for reform-induced spending that occurred when individuals were between the ages of 20 and 24. As further evidence of no effect for cohorts that were not exposed to reforms, the placebo estimates are in different directions for the various outcomes showing that there was no tendency toward improving or deteriorating outcomes among unexposed cohorts in districts that saw larger or smaller increases in school spending. These falsification tests support a causal interpretation of our estimates.

[2] Note that we do not use school spending prior to entering school as a placebo because there are numerous conditions under which spending at age 2 or 3 (prior to school entry) could have an effect on adult outcomes. For example an increase in school spending at age 2 could lead to the hiring of better teachers who may remain in the school for several years and still be at the school when the child enters. As such, one may find that spending at age 0 to 4 affects outcomes in adulthood even when our identification strategy is valid. Hence, spending prior to school entry is not a valid falsification test. Spending at age 20 -24 does not suffer from this problem so that it is a valid placebo. However, for this test we do include as a control the log of average spending during ages 0 to 4.

Appendix Table K1: *2SLS/IV Estimates of Court-Ordered School Finance Reform Induced Effects of Per-Pupil Spending on Long-Run Outcomes: Placebo Tests for Non-school Ages* (All children. All adult outcomes are measured between ages 20-45)

	Years of Education	Prob(High School Grad)	Ln(Wage)	Ln(Family Income)	Prob(Poverty)
	1	2	3	4	5
$Ln(PPE_d)_{(age\ 5\text{-}17)}$	3.2957**	0.6063***	0.8750***	1.0694***	-0.3739***
	(1.2963)	(0.1509)	(0.2029)	(0.2761)	(0.0976)
$Ln(PPE_d)_{(age\ 20\text{-}24)}$	-0.1700	-0.2878	-0.0705	-0.2499	-0.0959
	(2.5971)	(0.3954)	(0.4407)	(0.5954)	(0.2436)
Number of Individuals	15353	15353	13183	14730	14737
Number of Districts	1409	1409	1395	1414	1414

Robust standard errors in parentheses (clustered at school district level) *** p<0.01, ** p<0.05, * p<0.10

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative.

Models: The key treatment variable, $Ln(PPE_d)_{(age\ 5\text{-}17)}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$) and (the number of years of between the ages of 20 and 24 that occur after a court-ordered SFR) and (the number of years of between the ages of 20 and 24 that occur after a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$).

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends.

Appendix L: *Checking Robustness to Migration after Reforms*

Table L1: *2SLS/IV Estimates of Court-Ordered School Finance Reform Induced Effects of Per-Pupil Spending on Long-Run Outcomes: Using only Observations with Addresses Before the First Reform* (All children. All adult outcomes are measured between ages 20-45)

	Years of Education	Prob(High School Grad)	Ln(Wage)	Ln(Family Income)	Prob(Poverty)
	1	2	3	4	5
Ln(Spending)(age 5-17) × Low-income	4.1340***	0.9013***	1.0256***	1.7024***	-0.5945***
	(1.2642)	(0.2717)	(0.3051)	(0.3972)	(0.1449)
Ln(Spending)(age 5-17) × Non Poor	-0.0564	0.0387	0.5313	-0.0503	0.0398
	(1.3296)	(0.1863)	(0.4485)	(0.4179)	(0.0894)
Number of Individuals	15353	15353	13183	14730	14737
Number of School Districts	1409	1409	1395	1414	1414

Robust standard errors in parentheses (clustered at school district level) *** p<0.01, ** p<0.05, * p<0.10

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Observations for which the address data are obtained after the passage of the first court ordered SFR in their state are excluded from the analysis. Sampling weights are used so that the results are nationally representative.

Models: The key treatment variable, $\text{Ln}(\text{PPE}_d)_{(\text{age 5-17})}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of Spend_d).

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race) each interacted with linear cohort trends.

Appendix M: *Showing Similar Results for More Recent Versus Less Recently Educated Cohorts*

Even though we are careful to control for a variety of potentially confounding policies, one might worry that we have not accounted for some recent education reforms. For example, test-based accountability, charter schools and increased graduation requirements were all introduced in the late 1980s and early 1990s. While unlikely, if these recent reforms are correlated with the timing of court-mandated reforms, it could confound our estimates. To test for this, we estimate our 2SLS models of educational attainment interacting instrumented school spending with a dummy variable for those born between 1970 and 1985 (where earlier cohorts born between 1955 and 1969 serve as the comparison group). We focus on the education outcomes because differences in the effects on labor market outcomes across cohorts might be confounded with differences in economic conditions, returns to skills, and life cycle effects. If our effects are driven by other recent reforms, there should be no effect for the early cohorts, and the effect for more recent cohorts should be statistically significantly different from that of the older cohorts. We find that this is not the case. Instead, we find large, statistically significant effects for the older cohorts that are similar in magnitude to those for the full sample, and there is no statistically significant difference between the marginal effects for the older versus more recent cohorts. This evidence suggests that these other recent educational reforms are not driving our results and suggests little to no bias due to other more recent reforms.

Table M1: *2SLS/IV Estimates of School Spending on Education Outcomes for cohorts born 1955-69 vs 1970-85*

	Dependent variable:	
	Years of Education	Prob(High School Grad)
	1	2
Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$ Born '55-69	3.4965***	0.5618**
	(1.0885)	(0.2761)
Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$*Born '70-85	3.0772***	0.7265***
	(0.8524)	(0.1450)
F-test for differential spending effects by birth cohort (prob > F)	0.42	0.36
Number of Individuals	15,353	15,353
Number of Childhood Families	4,586	4,586
Number of School Districts	1,409	1,409

Robust standard errors in parentheses (clustered at school district level) *** p<0.01, ** p<0.05, * p<0.10

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Observations for which the address data are obtained after the passage of the first court ordered SFR in their state are excluded from the analysis. Sampling weights are used so that the results are nationally representative.

Models: The key treatment variable is, Ln(PPE_d)$_{(age\ 5\text{-}17)}$, is the natural log of average school-age per pupil spending. This is interacted with an indicator variable denoting whether the individual was born between 1955 and 1969 and an indicator variable denoting whether the individual was born between 1970 and 1985 (note: we do not include spending on its own so that there is no reference category). All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$) all interacted with an indicator variable denoting whether the individual was born between 1955 and 1969 or otherwise.

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race) each interacted with linear cohort trends.

Appendix N: *Validating the PSID Results Using Other Data*:

While the tests thus far show that our estimates are internally valid, readers might wonder how these patterns generalize to districts that are not included in the PSID. To address this, we replicated the analyses for high school graduation using the Common Core Data (CCD)—Local Education Agency Universe Survey and Non-Fiscal Survey Database—for all school districts in the US for available years 1987-2010 with the preferred research design. The model is the same as that employed for the PSID sample. However, it is important to note that there are numerous reasons to expect some differences between the results presented in the PSID and the CCD samples. First, because these data are at the district level rather than the individual student level and because the CCD data are based on the school district attended (rather than the school district of birth) any effects might reflect changes in school composition that occur as a result of changes in per-pupil spending associated with reforms. Second, the CCD data span a different time period from the sample analyzed in the PSID. While the PSID analysis is based on individuals who were of school age between 1960 and 1992, the CCD data span individuals who would have been school age between 1980 and 2008. Third, because outcomes in the CCD data are aggregated at the district level, the CCD results may suffer from aggregation bias. Fourth, while we have actual high school graduation data on the PSID, we only have proxies for the graduation rate in the CCD (the number of graduates divided by the number of 8^{th} graders four years prior). However, should the results be similar between the CCD data and the PSID sample, this robustness check would indicate that our findings are robust and generalizable.

Figure N1 presents the event-study graph for the number of high school graduates per 8^{th} grader (four year prior) among districts with a predicted spending increase. The pattern of results using the district level CCD data for 1987 through 2010 are similar to those using the individual level PSID data for 1976 through 2000. The 2SLS regression results in Table N1 using the CCD yield a point estimate of 35.48 (p-value<0.01) – indicating that increasing per-pupil funding by 10 percent over all 12 of a graduating cohort's school-age years would increase the number of graduates per 8^{th} grader by about 3.55 percentage points. We find it reassuring that the PSID and CCD point estimates are on a similar order of magnitude. However, we emphasize that the PSID estimates *are not* directly comparable to those from the CCD for all the reasons above. The important take-away is that the results from both datasets are qualitatively similar and point in the same direction – that money matters.

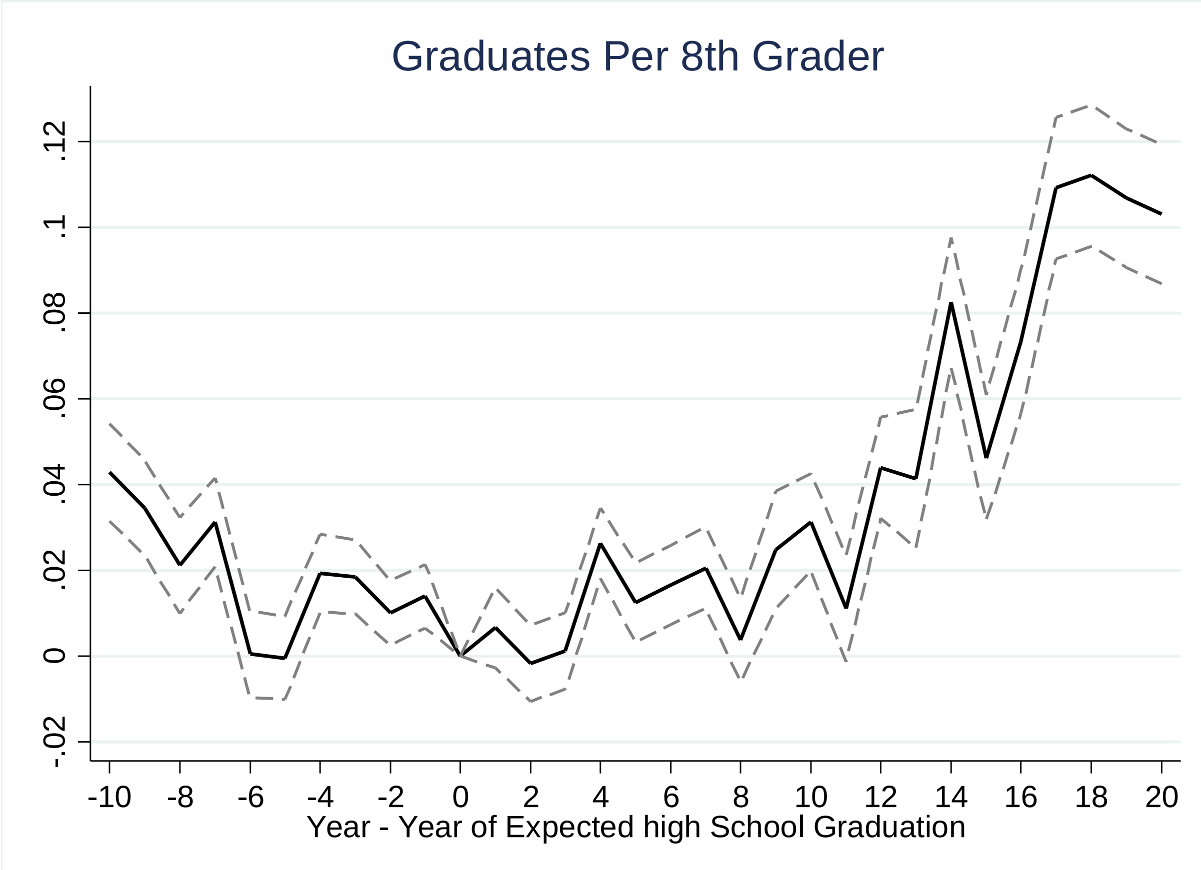
Because the CCD only has education outcomes, we also employ Census and American Community Survey (ACS) data for the same cohorts as those covered in the PSID. Unfortunately, these data do not link individuals to their school district during childhood, but only to their state of birth. While we cannot test for differences in outcomes using district-level per-pupil spending, we can test whether cohorts in states that saw increases in average school spending due to the passage of reforms have better outcomes than unexposed cohorts from the same state. It is important to note that using state-level increases in school spending as opposed to district level variation will likely lead to aggregation bias if the relationship between school spending and outcomes is non-linear. Accordingly, the relationship between state measures of school spending and outcomes will not be the same as the relationship between district measures of school spending and outcomes. It is also important to note that the Census data only include earnings for a single year. As shown previously, this will yield wage or earnings effects that are between one half and two-thirds the size found in the PSID. However, if there is a positive relationship between state measures of school spending and outcomes, it will lend further credibility to our results.

To mimic the PSID sample restriction, using a 3 percent sample from each decennial census and the ACS from 2000-2010, we only keep observations for individuals between the ages of 20 and 45 who were born between 1955 and 1985. This results in roughly 1.86 million observations. Because we do not have individuals linked to their childhood school districts we rely only on the variation in timing across cohorts within states. To illustrate that this variation is valid, we present event-study plots in Figure N2 of the effects of court-mandated school finance reforms on school-age per-pupil spending, the likelihood of not entering 12th grade (a proxy for dropout), the likelihood of high school graduation or GED receipt (not the same as high school graduation), and personal income. These figures only use variation in exposure to reforms (not variation in dosage). Because all variation is at the state level, the estimates for each event-study year is imprecise. However we present a linear fit for the pre-reform years and the post-reform years along with the 95 percent confidence intervals.

For all outcomes, the event-study plots show (a) minimal pre-reform differences in cohort trends between reform and non-reform states, (b) clear improvement in all outcomes after the passage of a court-ordered school finance reform, and (c) improved outcomes with increased years of exposure. Given that this is only using state-level variation in the timing, these patterns are reassuring. To quantify these patterns, we estimate 2SLS models where we predict outcomes as a function of school spending, conditional on state fixed effects, census region trends, gender, and race-specific cohort trends, and age. We instrument for school spending with indicator variables for the number of years of school-age exposure to reforms interacted with the median predicted spending change ($Spend_d$) for that state. This instruments exploit variation across cohorts within states and also variation across states in average district-level dosage. Results are presented in Table N2. The 2SLS results in Columns 4 through 6 show that a 10 percent increase in median state per-pupil spending increases the likelihood of graduating from high school or earning a GED by 0.978 percentage points (p-value<0.05), reduces the likelihood of not entering 12th grade (a proxy for dropout) by 0.85 percentage points (p-value<0.05), and increases individual income by $1,715 (p-value<0.10). Relative to base income levels, this represents a 5.2 percent earnings increase. Recall that the Census results are based on a single observation per person, so that the estimates will be between one-half and two-thirds of the effect on permanent earnings. As such, the Census estimates imply real earnings effects of a 10 percent increase in median state per-pupil spending of between about 7 and 11 percent. Even though, one might expect very different results using state level spending changes and district level spending changes, the implied earnings effects using the Census data are well aligned with the PSID estimate of 7.5 percent.

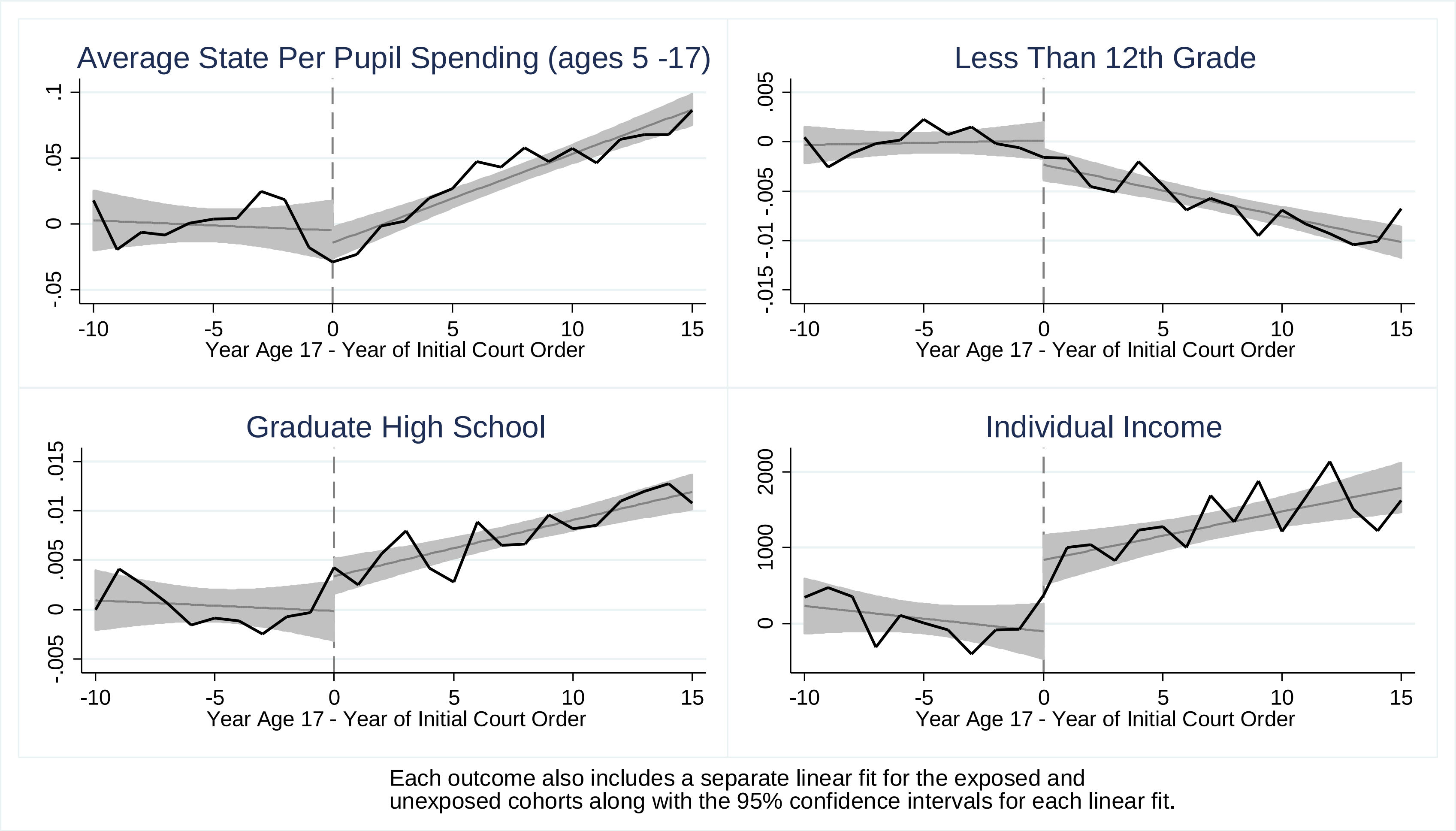
We have replicated patterns of economically meaningful effects of school spending on adult outcomes across different datasets, for different cohorts, using different sources of variation, and using data at different levels of aggregation. Given the inherent differences between the various models, unsurprisingly, the point estimates are not identical. However, across all datasets, cohorts, sources of variation, and levels of aggregation we are able to rule out small effects and zero effects—providing compelling evidence that exogenous spending increases improve children's long-run outcomes.

Figure N1: *Effect of the Initial Court-ordered Reforms on High School Graduation Rates for Districts with Predicted Increases (w/ 90% CI): Using the CCD and Full Population of Districts from 1987 – 2010.*



Data: CCD Data from 1987-2010 for 15,353 districts
Models: The event-study plot is based on indicator variables for the number of years of exposure to a court ordered SFR interacted with an indicator for whether the district was in the top quartile of the state distribution of per-pupil spending in 1972. Results are based on non-parametric event-study models that include school district fixed effects, race×census division×birth cohort fixed effects and additional controls.
Additional controls: census division × year fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) each interacted with linear time trends.

Figure N2: *Effect of Court-Mandated Reforms on Spending and Outcomes: IPUMS Census Data*



Data: Individual Census IPUMS Data (1970, 1980, 1990, and 2000-2012) matched with the timing of court-mandated reforms by date of birth and state of birth. Models: The event-study plot is based on indicator variables for the number of school-age years of exposure to a court ordered SFR. Results are based on event-study models that include: state of birth fixed effects, year of birth effects, age and age squared, and gender and ethnicity interacted with a linear cohort trend. The figure plots the estimated effects of each year of exposure to an initial school finance reform for each outcome. We also include the linear fit for the exposed and unexposed cohorts along with the 95 percent CI for each linear fit.

Table N1: *Effects on high School Graduates per 100 8th Graders and Dropout Rate in the CCD*

	1	2	3	4
	CCD Data			
	Graduates Per 100 8th Graders		Dropout Rate (0-100)	
Mean of Dependent Variable	82.93		3.55	
	OLS	2SLS	OLS	2SLS
Ln ($PPE_{(age\ 5\text{-}17)}$)	-0.325	35.48***	-1.069***	-21.55***
	(0.714)	(7.666)	(0.236)	(2.228)
Observations	125,659		111,570	

Robust standard errors in parenthesis adjusted for clustering at the school district level. *** p<0.01, ** p<0.05, * p<0.1

Data: CCD Data (1987-2013).

Models: The key treatment variable, $Ln(PPE_d)_{(age\ 5\text{-}17)}$, is the natural log of average school-age per pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below.

2SLS: The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$). The first stage F-statistic is greater than 20 in all models.

Additional controls: census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, roll-out of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends.

Table N2: *Effects on High School Equivalent Holder and Personal Income in the IPUMS Census Data*

	1	2	3	4	5	6
	OLS			2SLS: Variation Across Cohorts within States		
	Graduate High School or GED	Less Than 12th Grade	Total Individual Income	Graduate High School or GED	Less Than 12th Grade	Total Individual Income
$\text{Ln(PPE)}_{\text{(age 5-17)}}$	0.0152**	-0.0123**	1,206	0.0978**	-0.0854**	17,154*
	(0.00640)	(0.00555)	(1,116)	(0.0454)	(0.0429)	(10,214)
Mean of Dep. Variable	0.989	0.074	32,558	0.989	0.074	32,558
First Stage F-statistic	-	-	-	193.38	193.38	193.38
Observations	1,862,165	1,862,165	1,860,118	1,862,165	1,862,165	1,860,118

Robust standard errors in parenthesis adjusted for clustering at the state level.

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

Data: IPUMS Data for the following years (1970, 1980, 1990, 2000 – 2011).

Models: The key treatment variable, $\text{Ln(PPE}_d)_{\text{(age 5-17)}}$, is the natural log of average school-age per pupil spending. All models include state fixed effects, year of birth by race fixed effects, race-by-year linear trends, gender fixed effects, and age indicators.

2SLS: The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the median value of $Spend_d$ in the state). The first stage F-statistic is greater than 10 in all models.