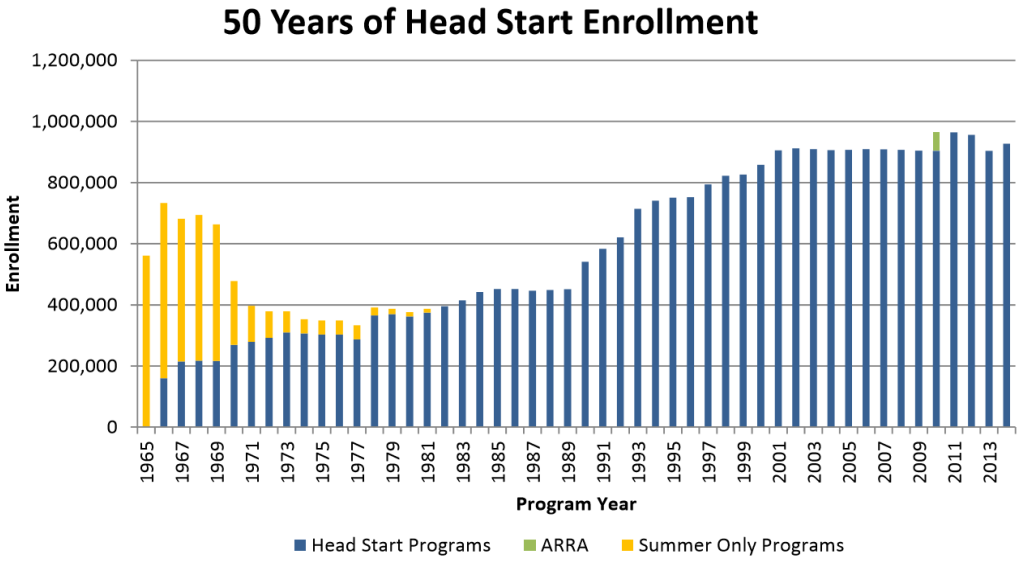
ONLINE APPENDIX

REDUCING INEQUALITY THROUGH DYNAMIC COMPLEMENTARITY: EVIDENCE FROM HEAD START AND PUBLIC SCHOOL SPENDING

RUCKER C. JOHNSON AND **C. KIRABO JACKSON**
UC-BERKELEY & NBER NORTHWESTERN UNIV & NBER

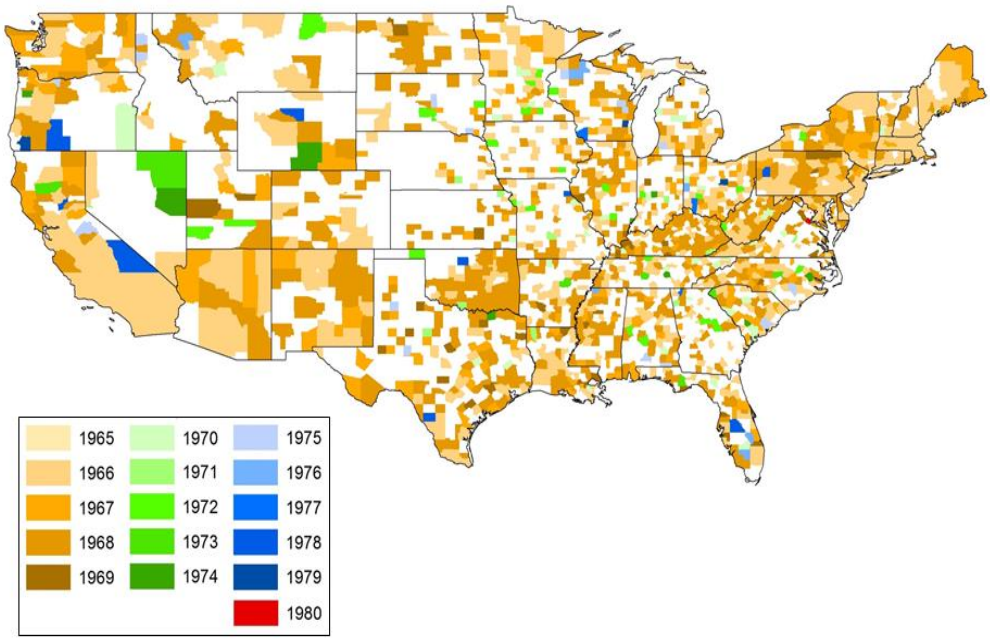
Appendix A
Additional Tables

Figure A1:
National Head Start Enrollment over Time



Note: Chart is pasted directly from the Head Start Fact Sheet (link: https://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2015-hs-program-factsheet.html)

Figure A2:
Year of Establishment of First Head Start Center by County



Note: Based on authors' calculations and data collections as described in Appendix D.

Appendix B

Spending per enrollee versus spending per eligible

The expansion of Head Start involved both increases in the number of enrolled children and increases in spending per enrolled child. Head Start spending per enrollee increases do not capture increases in the total number of children affected by Head Start, so that spending per poor four-year-old in the county is a more appropriate measure. To illustrate this point, we collected data on Head Start spending per enrollee and Head Start spending per poor 4-year old at the state level between 2003 and 2014 (years for which both sets of data are available). Using within-state changes in spending over time, a 10% increase in spending per poor four-year-old is associated with only a 0.243%.

Table B1:

Relationship between Spending per Enrollee, Spending per Poor 4-Year-Old and Enrollment (at state-year level)

	1	2	3	4	5	6	7	8
	Spending per Enrollee		Log of Spending per Enrollee		Log of Head Start Enrollment		Share of Income Eligible four-	
Spending per poor 4-year-old	0.0174	0.0379*						
	[0.0359]	[0.0143]						
Log Spending per poor 4-year-old			-0.0192	0.0243*	0.0810+	0.1207**	0.648**	
			[0.0271]	[0.0093]	[0.0482]	[0.0438]	[0.0506]	
Log Spending per Enrollee								-0.130
								[0.208]
Year FX	N	Y	N	Y	N	Y	Y	Y
State FX	Y	Y	Y	Y	Y	Y	Y	Y
Observations	612	612	612	612	612	612	612	612
R-squared	0.627	0.921	0.638	0.927	0.991	0.995	0.930	0.631

Robust standard errors in brackets adjusted for clustering at the state level

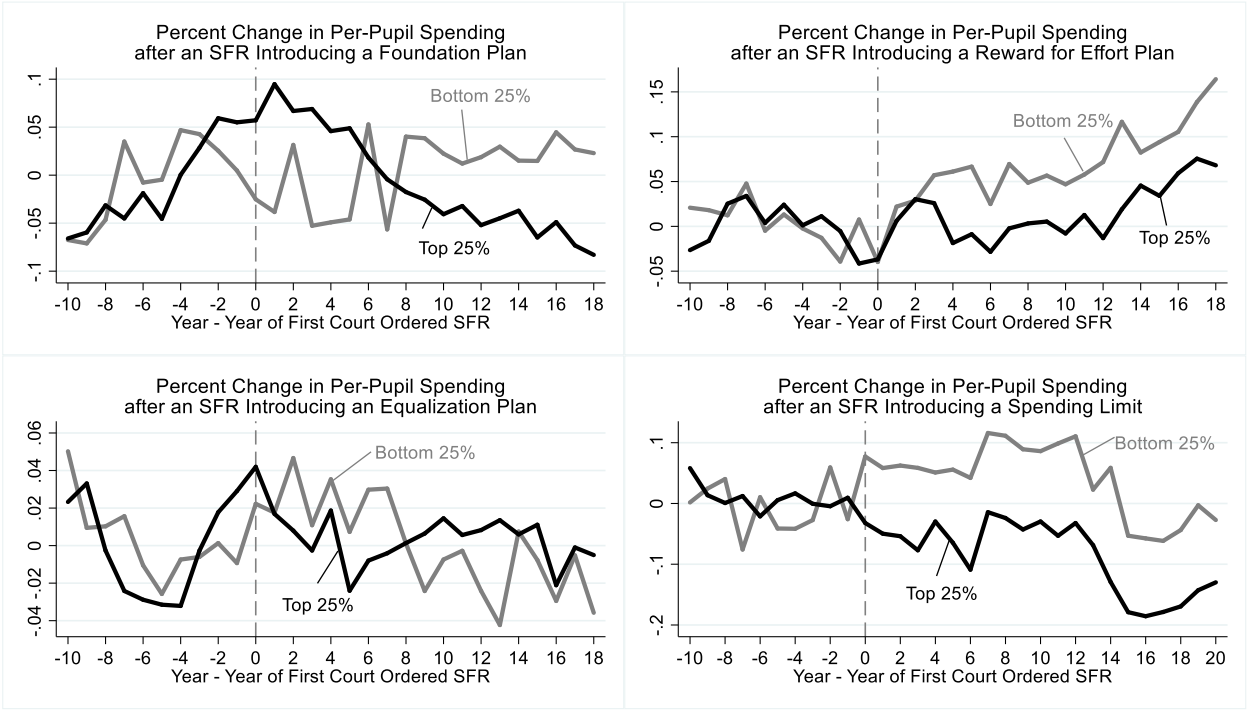
** p<0.01, * p<0.05, + p<0.1

Notes: State year level data on total federal Head Start spending and total Head Start enrollment is obtained from the Head Start Facts fiscal years reports 1999 through 2015. Data on the number of poor four-year-olds in the state in each year is obtained from Integrated Public Use Microdata Series (IPUMS) microdata that preserves and harmonizes decennial censuses from 1790 to 2010 and American Community Surveys (ACS).

Appendix C

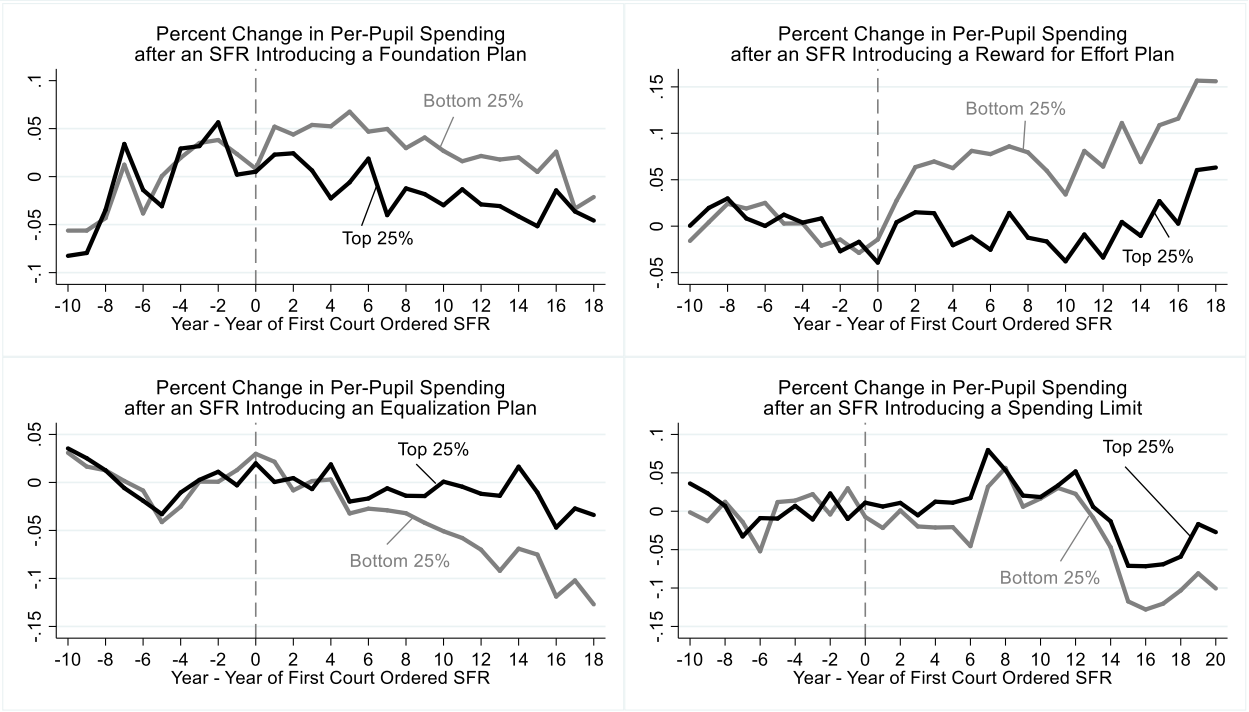
To illustrate how the introduction of different formula types affected districts by pre-reform income and spending levels, we replicate the analysis in Jackson Johnson and Persico (2016). Figures C1 and C2 present event-study plots of the natural log of per-pupil spending at the district level (after removing both district and year fixed effects). Year 0 is the first year of the first court order in the state, year "-5" is five years before the first court order, and year "5" is five years after the initial court order. For each court order, we link all formula changes that occurred within three years to that court-ordered SFR. Figure C1 shows the evolution of per-pupil spending for districts in the bottom and top quartiles of per-pupil spending in 1972 (the year preceding the first court-ordered SFR) after court orders that led to the implementation of different kinds of funding formula plans. Figure C2 presents similar plots for districts in the top and bottom quartiles of the state income distribution in 1969. Figures C1 and C2 show that court-ordered SFRs that lead to the implementation of different funding formulas had different effects on districts by pre-reform income and spending levels.

Figure C1: *Effect of Formula Type on District Per-Pupil Spending by District Spending in 1972*



Data: The sample includes all school districts in the United States between the years of 1967 and 1999. The sample is made up of 324,321 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% 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 C2: *Effect of Formula Type on District Per-pupil Spending by District Income in 1969*



Data: The sample includes all school districts in the United States between the years of 1967 and 1999. The sample is made up of 324,321 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% 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.

Appendix D: *Panel Study of Income Dynamics (PSID, 1968-2015)*

The PSID began interviewing a national probability sample of families in 1968. These families were re-interviewed each year through 1997, when interviewing became biennial. All persons in PSID families in 1968 have the PSID "gene," 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. Moreover, the genealogical design implies that the PSID sample today includes numerous adult sibling groupings who have been members of PSID-interviewed families for more than four decades. 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.

The PSID maintains high wave-to-wave response rates of 95-98%. Studies have concluded that the PSID sample of heads and wives remains representative of the national sample of adults (Fitzgerald, Gottschalk, Moffitt, 1998a,b; Becketti et al, 1988). Additionally, we perform a supplementary analysis of sample attrition in the PSID, and find no evidence of selective attrition among our study sample (Appendix Table D1). In particular, among original sample children, baseline 1968 family and county characteristics do not jointly significantly predict the likelihood of attrition or the likelihood of being observed as an adult.

The share of individuals potentially exposed to Head Start expenditures at age 4 increases significantly with birth year over the 1950-1976 birth cohorts analyzed in the PSID sample. Two-thirds of the sample grew up in a state that was subject to a court-mandated SFR between 1971 and 2000 (the first court order was in 1971).

Matching PSID Individuals to their Childhood School Districts

We use the confidential restricted-use geocode PSID data that includes census block identifiers that correspond with childhood respondent addresses. We match respondent earliest childhood residential location (typically, 1968) to school districts via the combination of GIS mapping methods and school-to-census tract relationship files. 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. 1970 street addresses for schools are obtained from the Elementary and Secondary General Information System (ELSEGIS). Using GIS software, we locate these schools using the 2000 census electronic road maps (http://www.esri.com/data/download/census2000_tigerline/). We use a crosswalk of census tract identifiers across 1970/1980/1990/2000/2010 censuses (since the definitions of neighborhoods change over time), and 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.

Table D1: *PSID Analysis Tests of Sample Attrition***.**

	Dependent variable:		
	Probability(original sample child observed in adulthood)		
	All	**Poor Children**	**Non-Poor Children**
	(1)	(2)	(3)
1968 Family & County Characteristics:			
Black (ref cat: white)	0.000241	-0.004006	0.003434
	(0.01807)	(0.02313)	(0.03497)
Family income-to-needs ratio	0.001694	0.007748	-0.0001174
	(0.006366)	(0.02573)	(0.008722)
Female-headed household	-0.03048	-0.01692	-0.06152
	(0.01874)	(0.02210)	(0.04078)
Number of children	0.006430	0.007830	0.008396
	(0.004636)	(0.006050)	(0.008612)
Parental education (ref cat: high school grad):			
High school dropout	-0.01927	-0.01291	-0.02987
	(0.01659)	(0.02264)	(0.02559)
Attended college	0.01494	-0.03191	0.03238
	(0.02214)	(0.04690)	(0.02446)
Household annual food expenditures	-0.0002157	-0.0002115	-0.0003181
	(0.0001734)	(0.0003127)	(0.0002228)
Parental expectations for achievement, index	-0.002894	-0.001608	-0.003156
	(0.003762)	(0.005337)	(0.005312)
County unemployment rate	0.002809	0.01084	-0.008268
	(0.008466)	(0.01264)	(0.01120)
County public assistance expenditures per capita	0.0007209	0.001196	0.0004415
	(0.002912)	(0.004356)	(0.004037)
Region (ref cat: South):			
Northeast	0.0002421	-0.02104	0.03062
	(0.01943)	(0.02879)	(0.02690)
Midwest	0.0004922	-0.008726	0.01732
	(0.01833)	(0.02709)	(0.02628)
West	-0.004869	-0.03543	0.02812
	(0.02173)	(0.03525)	(0.02809)
F-test of joint significance (p-value)	F-stat=0.83;	F-stat=0.42;	F-stat=1.12;
Proportion of original sample children observed in adulthood	0.766	0.753	0.766

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2015): Analysis sample includes PSID original sample children born 1950-76. 76.6% of these children have been followed into adulthood, and are included in analysis sample in main results presented in Tables 2 and 3. This Table shows no evidence of selective attrition based on 1968 childhood family and county characteristics. (p-value of F-test of joint significance of vars = 0.8301).

Appendix E

County-Level Federal Outlays for Head Start and Title I, 1965-1980

Our collection of Head Start data follows Johnson (2015). County-year federal outlays for Head Start and Title I ESEA were computed using county-level federal outlays data acquired from the National Archives and Records Administration (NARA) for fiscal years 1965 through 1980, along with ICPSR Study #6029 (for fiscal years 1976 to 1980). Information was culled from NARA records by searching program titles and program codes. We identified the pool of grants for Head Start from the NARA records, which included string searches on Head Start grant titles. For most records, Head Start programs are listed by community and funding amounts, and information on the "stock" of programs at a particular time allows verification of the accuracy of grant "flows". Likewise, we identified the pool of grants for Title I/ESEA outlays from the NARA records by using program titles and program codes over this period. The county-year federal Head Start and Title I outlays were converted into 2000 dollars using the CPI-U deflator.

County-level information on Community Action Program (CAP) Grants and grantees on federal CAP grants is derived from the NARA microdata (Community Services Administration 1981). These data files document neighborhood and community-based poverty programs as funded by CAP and CAP grant-action data include data on the target population of grant proposals. These records are structured as two data files spanning 1965 through 1980. One data set is observed at the level of individual grant actions; the other dataset records data on the organizations receiving grants. The combined data include information on any "action" on a grant (when it is recorded, extended, renewed, or terminated), dates associated with these actions, and some information about the funded project. We use the county-level geographical identifiers from the grantee data and grant-action file, which include the name and county of designated grantee and county where the services are provided in most cases. We aggregate these amounts by the fiscal year of disbursement and county of service delivery. These amounts have been verified by state against information printed in OEO annual reports (Office of Economic Opportunity, 1965–1968).

We compared our calculated county-level federal outlays for Head Start with those reported in Ludwig and Miller (2007) for fiscal years 1968 and 1972, and Elizabeth Cascio (2009) for 1976-80, and in each case our numbers line up with those used by these authors (who generously shared their data for comparison). Our county-level panel of Head Start spending though spans a much longer time period than used in previous studies. We compared spending totals calculated from the county-level files to published data at the federal level and state level (where available) to assess the validity of the county-level data. Following Cascio (2009), we compared the state-level Head Start outlays calculated in our data to those reported in Jones (1979) for fiscal years 1970 through 1977, and the correlation coefficient was above 0.975 in all fiscal years except 1974, where Mississippi was an obvious outlier. We, therefore, dropped all fiscal years for Mississippi for the Head Start analysis because the reporting of federal outlays for that state at the county-level had some obvious errors and were poorly documented.

We then assembled population counts of the number of 4-year olds and the number of school-age children ages 5-17 in every US county, respectively, using the Surveillance, Epidemiology, & End

Results (SEERS) program data spanning the period 1965 through 1980. The county-year federal outlays for Head Start and Title I ESEA were combined with both the county-year population counts of the number of 4-year olds and number of children ages 5-17, and the 1970 county-level poverty rates among children (and non-elderly persons) in order to construct our measures of county-level Head Start spending per poor 4-year-old and county-level Title I (ESEA) spending per-pupil, for 1965 through 1980.[1] Note that the SEERS data are not broken down by poverty and age. As such, we obtain the 1970 county-level child poverty rate via the 1970 Census (ICPSR) data and multiply this by the county-level number of 4-year olds, which together provides an accurate estimate of the number of poor 4-year olds in each county (assuming county-level child poverty rates do not differ greatly by child age).

District-Level K-12 School Spending 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

[1] References for the data appendix:

- Cascio, Elizabeth (2009). "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools". NBER Working Paper No. 14951.
- Johnson, Rucker C. (2015). "Follow the Money: School Spending from Title I to Adult Earnings". Special edited volume, ESEA at 50, published in *The Russell Sage Foundation Journal of the Social Sciences*.
- Jones, Jean Yavis (1979). "The Head Start Program – History, Legislation, Issues and Funding 1964-1978". Washington, D.C.: Congressional Research Service. Report 79-14 EPW.
- Ludwig, Jens and Douglas L. Miller (2007). "Does Head Start Improve a Children's Life Chances? Evidence from a Regression Discontinuity Design*." Quarterly Journal of Economics* 122(1): 159-208.
- Office of Economic Opportunity. Annual Reports. Washington, DC: GPO, 1965–1968.
- U.S. Bureau of the Census. Census of Population Supplementary Report. Poverty Status in 1969 and 1959 of Persons and Families, for States, SMSA's, Central Cities and Counties: 1970 and 1960.

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.

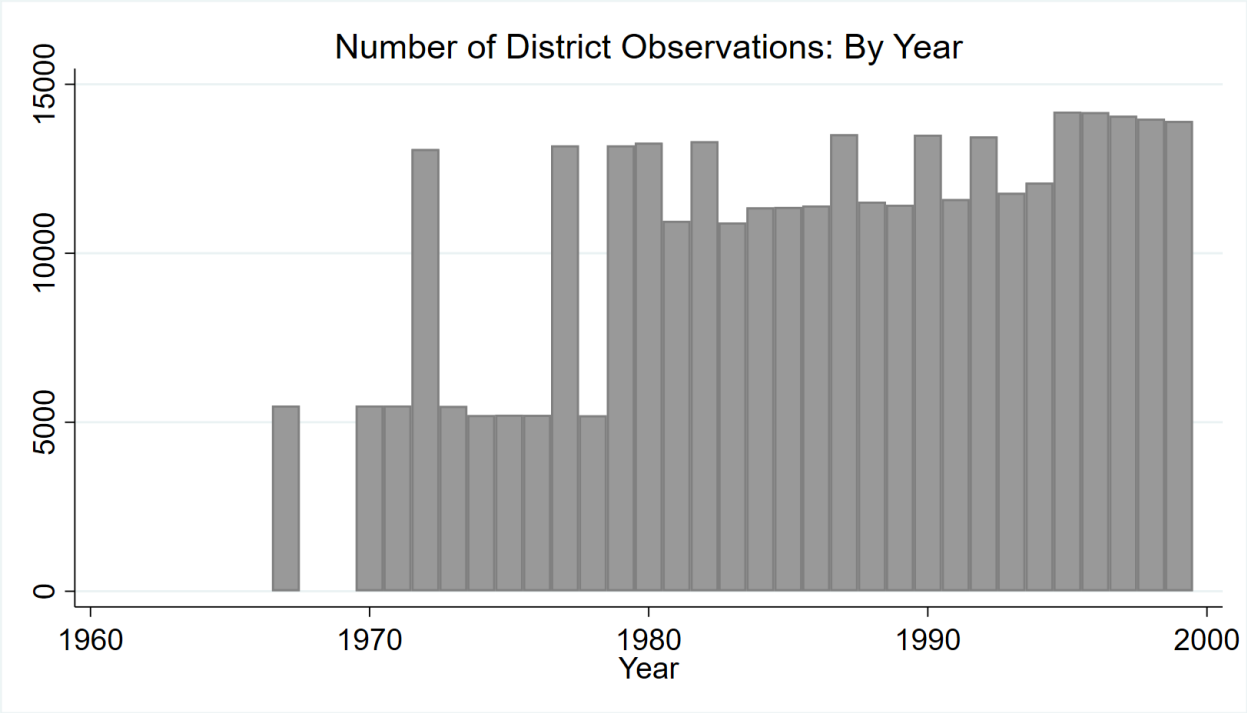
In addition to standardizing the data, the Census Bureau has corrected a number of errors in the INDFIN database that were previously in other sources of data. For example, for fiscal years 1974, 1975, 1976 and 1978 the school district enrollment data that had previously been released were useless (either 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 2000. 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 E1 below shows the number of district observations in our data for each year. While

the coverage of the data we use is arguably better than that used previously, it is not perfect. As shown in Appendix Figure E1, for years, 1967, 1970, 1971, 1973, 1974, 1975, 1976, and 1978 only about 40% of districts are present (often larger districts). After 1979 almost all districts are included.

Figure E1: *The number of district observations for each year.*



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 United 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).

Data on Other Policies and Additional Controls

The data we use for other controls 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; county-level Title I/ESEA spending (NARA); the comprehensive case inventory of court litigation regarding school desegregation over the 1955-1990 period (American Communities Project), and major plan implementation dates in large districts (compiled by Welch/Light); and American Hospital Association's Annual Survey of Hospitals (1946-1990) and the Centers for Medicare Provider of Service data files (dating back to 1960s) to identify the precise date in which a Medicare-certified hospital was established in each county of the US (an accurate marker for hospital desegregation compliance).

Appendix F
Predicting Dosage

The prediction of $dose_d$ is obtained in two steps. We discuss each step in turn below.

Step 1

First, using district-by-birth-cohort data for the full universe of districts (not only those represented in the PSID), we use flexible Difference in Difference regression models to predict how school spending in each district responded to the passage of a court-ordered reform based on (a) the type of reform introduced after the court order interacted with the district's school spending levels prior to reforms, and (b) the type of reform introduced after the court order interacted with the district's income level prior to reforms.

To do this, we estimate [F1] where all common variables are defined as in [4] and [5]. In [F1], *T* is event time and is the year an individual turned 17 minus the years of the first court-ordered SFR in their state of birth. Accordingly, *T* is 0 for those who turned 17 the years of a SFR and are essentially not exposed, it is -2 for those who turned 19 during the year of an SFR so that they graduated high school 2 years before the SFR, and T would be 5 for individuals who turned 17 f years after the first SFR in their state of birth. This exposure measure varies at the state birth-cohort level and goes from -20 (those who were age 17 twenty years before the state's first court-ordered SFR) to 12 (for those who were ages 5 and younger the year of the state's court-ordered SFR). In [F1], $I_{F,d}$ is an indicator for the type of reform (*F*) (i.e−*foundation plans, spending limits, reward for effort plans, equalization plans, and equity cases*) introduced by the court order in the state containing district *d*, $Q_{ppe72,d}$ is the quartile of district *d* in the state distribution of per-pupil spending in 1972, and $Q_{inc69,d}$ is the quartile of district *d* in the state distribution of median income in 1969.

[F1] $$\ln(PPE_{5-17})_{idb} = \sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\right)\cdot\alpha_{T,Q_{ppe}} + \sum_{F=1}^{5}\sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{inc69,d}=Q_{ppe}}\times I_{F,d}\right)\cdot\alpha_{T,Q_{inc},F} + \Pi C_{idb} + \theta_{d3} + \theta_{b3} + \varphi_{idb}.$$

The variables $\sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\right)$ are the set of interactions between the quartile of district *d* in the state distribution of per-pupil spending in 1972, and exposure to an SFR. Accordingly, 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, $\sum_{F=1}^{5}\sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\times I_{F,d}\right)$ are the set of interactions between the type of reform, the quartile of district *d* in the state distribution of median income in 1969, and exposure to an SFR. Accordingly, 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.

Step 2

In the second step, we take the estimates from estimation of [F1] to summarize how a given districts per pupil spending is likely to change after the introduction of a courted ordered SFR in their state. That is, for each district we use the predicted spending change (based on reform type implemented by the state and district spending and district income levels prior to reforms) for those

who were between the ages of 10 and 15 in the year of the initial court-ordered SFR (i.e., those six cohorts exposed to an SFR for between 3 and 8 years). To assuage any concerns that this age range choice is arbitrary, note that our results are similar when using other ages such as ages 10 to 17 or 5 to 17.[2] Formally, our predicted district-specific dose effect based on [F1] is

[F2] $$\widehat{\text{dose}}_d = \left[\sum_{Q_{ppe=1}}^{4} \sum_{T=3}^{8} \left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}}\right) \cdot \hat{\alpha}_{T,Q_{ppe}} + \sum_{F=1}^{5} \sum_{Q_{ppe=1}}^{4} \sum_{T=3}^{8} \left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}} \times I_{F,d}\right) \cdot \hat{\alpha}_{T,Q_{inc},F}\right] / 6$$

By using the predicted values, $\widehat{\text{dose}}_d$, from [F2] from the full universe of school districts as an instrument in a 2SLS regression on the PSID sample, we implement a two-sample instrumental variables (2S-2SLS) strategy where our excluded instruments are the exposure indicator variables interacted with a function of the reform type implemented by the state, the district income level prior to reforms, and the spending level of the district prior to reforms.[3] This approach captures meaningful variation in K12 spending due to the reforms but removes any variation in spending that is determined by local factors that also influence outcomes.

NOTE: We estimate our main models excluding $\widehat{\text{dose}}_d$, (i.e. using only variation due to exposure to an SFR) and the results are very similar. See appendix Table H2.

[2] We chose this age range because it included enough years (i.e., 5) to not be sensitive to random fluctuations in the high frequency data, and because it occurred relatively soon after the passage of a court-ordered reform (these exclude the first two years following a court order as there was typically a two-year delay in legislative implementation of SFRs following a court order, with limited spending changes in the year immediately after.

[3] The two-sample 2SLS estimator was popularized by Angrist and Krueger (1992) and has been used successfully in several other empirical settings (e.g. (e.g., Bjorklund and Jantti, 1997; Currie and Yelowitz, 2000; Dee and Evans, 2003; Borjas, 2004).

Appendix G

Estimated Effects on Head Start Participation

To get a sense of how our spending increases relate to changes in the Head Start participation margin, we used changes in national Head Start enrollment over time. However, given that Garces, Currie, and Thomas (2002) employ data on Head Start participation reported by PSID respondents, it is important to discuss the implied participation effects using these data. The data on Head Start participation used in Garces, Currie, and Thomas (2002) are imperfect in important ways. First, the data are retrospective data collected in the 1995 survey wave based on questions that asked adults about their early childhood experiences and whether they had ever participated in a Head Start program. Even though Garces, Currie, and Thomas (2002) present some evidence that any recall bias in these data may not be severe, we are reluctant to trust these data when there are other alternatives. Potential recall bias may be particularly problematic for Head Start participation during the ramp-up period during which most of our variation is derived. This is due to the fact that the largest increases in Head Start enrollment occurred between 1965 and 1970 in the summer-only programs, which were largely phased out 1970 onwards. As such, the large increases in Head Start participation (much of which were in the summer-only programs) between 1965 and 1970 are not reliably recorded in the participation survey responses reported in the PSID. As a result, relating increases in Head Start spending to retrospectively reported Head Start participation in the PSID might drastically understate the effects of Head Start spending on enrollment in the program.

Having discussed the limitations of using the reported Head Start enrollment in the PSID to infer the effects of spending on enrollment, it is helpful to show what estimates these data yield. To explicitly model the relationship between increased spending on Head Start and the participation of low-income children in Head Start (using the self-reports from the PSID) we estimate conditional logit models. We predict Head Start participation using Head Start spending per poor four-year-old in the county while controlling for race/ethnicity and conditioning on the childhood county. We exclude controls for cohort trends because, by definition, such trends are zero before rollout. To allow for ease of interpretation, we report the average marginal treatment effects based on the conditional logit estimates. The marginal effects are presented in Table F1 for children from poor families.

The point estimates reveal that, for the poorest children, increasing Head Start spending by $1,000 per poor four-year-old would increase the likelihood of reporting enrollment in Head Start by 18.02 percentage points. This implies that for the average county that spends $4,320 per poor four-year-old, Head Start participation is estimated to increase by 18.2*4.32=78.6 percentage points. We also estimate the effect of Head Start rollout (e.i. if a Head Start center was available in the individual's county of birth when they were 4 years old) on participation. The marginal effect from the conditional logit model is 0.797. This is almost identical to the implied effect from the spending specification and suggests that rollout increased Head Start participation by roughly 80 percentage points among children from poor families. Because the conditional logit model requires that there be some variation in the outcome within each county, we cannot run the conditional logit models on the non-poor population because the vast majority of counties do not have any non-poor children in Head Start.

This implied participation effect at rollout of about 80 percentage points for poor children is very similar to our assumed participation effect of 75% for any Head Start program, somewhat larger than our assumed participation effect of about 63% for full-time Head Start, and roughly the same order of magnitude as both. Because these estimates are on a similar order of magnitude as those computed based on national data, we are confident that our preferred estimates of the participation margin from the national data are reasonably accurate.

Table G1:
Conditional Logit Estimates of the Effects of Head Start Spending on Head Start Participation of Poor Children

	1	2
	Prob(Attend Head Start)	Prob(Attend Head Start)
	Conditional Marginal Effects, evaluated at means	
County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s)	0.1802***	
	(0.0205)	
Head Start Center(age 4)		0.7972***
		(0.0500)
Number of Children	4,651	4,651
Number of Childhood Families	1,909	1,909
Number of School Districts	631	631
Number of Childhood Counties	448	448

Robust standard errors in parentheses (clustered at childhood state)
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes PSID individuals born 1950-1976 who were followed through the 1995 survey IW. Child-specific pre-K attendance & Head Start program participation information collected retrospectively in 1995 survey IW. Poor children are those whose parents were in the bottom quartile of the income distribution (approximately 80% of whom were below the poverty line).
Models: Results are based on models that include school district fixed effects and controls for race/ethnicity.

Appendix H

Robustness Checks and Tests of Validity

Because one of the parameters of interest is the marginal effect of the interaction between Head Start Spending and public K12 spending, it is important for us to establish that the variation we use in each of these is exogenous and will yield causal relationships. Here we present a series of empirical tests that support the validity of each source of variation.

Head Start Spending Effects by Child Age: *No confounding policies*. As a falsification/placebo test, we investigate the effects of Head Start spending increases by the child's age at which these increases occur. If the results are consistent with a causal interpretation of Head Start spending, then we would expect to find significant effects of that spending only for children who are age-eligible (age 4), and not for children who were already school-age at the time of the spending increase. Furthermore, even though our models control for a variety of other policies and we find no Head Start effects on non-poor children, one may still worry that the timing of Head Start rollout or the timing of SFRs coincided with other policies that also improved adult outcomes. One test of this would be to determine whether the effects of the spending increases are experienced only among those who were of the appropriate age. If counties or districts adopted other policies to improve outcomes for low-income children (that were not targeted to the exact same age range as that in question) one would observe improvements for other age ranges also. To test this for Head Start spending, we estimated the marginal effect of the level of Head Start spending that prevailed when the individual was different ages. To test whether Head Start spending at other ages predicts student outcomes, conditional on Head Start spending at age 4, we estimate the following regressions where all variables are as defined in [4] and [6].

[H1] $$Y_{icb} = \beta_w \cdot HS_{cb}^{age\,W} + \beta \cdot HS_{cb}^{age\,4} + \gamma \cdot C_{icb} + \theta_c + \tau_b + \varepsilon_{icb}.$$

We estimate models such as [H1] where we include our regressor of interest (β_w), the marginal effect of Head Start spending at age W on individual outcomes, conditional on the effect of Head Start spending at age 4. In principle, one should see that Head Start sending per poor four-year old has effects when the individual was four years old but not at other ages. This is exactly what we find across every one of the adult outcomes of poor children we analyze. In Figure 6 and Figure H1, we plot the marginal effect of Head Start spending by age conditional on spending at age 4. Note that the estimated effect for age 4 is not conditional on spending at other ages. However, the marginal effect of spending at age 4 is largely the same in models that include spending at other ages. The figures all show that increases in the Head Start spending level that prevailed when the individual was four years old are associated with significantly improved adult outcomes while the corresponding spending level at ineligible ages (1-3;5-10) are not.

Even though we instrument for K12 spending levels, it is important to establish that the identifying variation we use is valid. If the spending increases we exploit operate through improved K12 education, one should see improvement for those who were between the ages of 5 and 17 when there was a school finance reform, but no effect for individuals from the same districts who were 18 or older at the time. Figure 5 shows that only those individuals who were of school-going age at the time of a reform-induced spending increase experience improved outcomes. These figures also reveal that outcomes in districts that saw increases in K12 spending were not on a positive or negative trajectory – indicating that the timing of the SFR was exogenous to the underlying trends in outcomes in affected districts. To show this more formally, we estimate models that instrument for K12 spending in an individual's childhood districts when they were

between the ages of 20 and 24. Results are in appendix Table H1. If the effects are real, we should see effects for reform-induced spending increases when an individual was between the ages of 5 and 17 but not for increases that occurred when an individual was between the ages of 20 and 24. As in Jackson, Johnson, and Persico (2016), K12 spending levels between ages 20 to 24 have no effect on outcomes.

Robustness to using exposure variation to SFRs only: Our approach to estimating isolating the causal effect of school spending on student outcomes is somewhat complicated. While our approach makes the best use of all plausible exogenous variation in school spending due to the passage of a court-ordered school finance reform, it is helpful that our results are robust to using a simpler approach. Specifically, one may worry that our dosage predictor ($\widehat{dose}_d$) may be biased, and that because it is estimated in a first stage, we may understate the underlying noise in our final estimates. To address both these concerns directly, we estimate models that only use variation in SFR exposure for identification *and do not use any variation due to dosage*. Specifically, we use the within-county DiD variation in Head Start spending ($HS_{icb}^{age\,4}$), instrument for the natural log of public K12 spending, ppe_{idb}^{5-17}, with ($SFRExp_{idb}$), and instrument for, INT_{idb}, the interaction between Head Start and K12 spending, with ($HS_{icb}^{age\,4} \times SFRExp_{idb}$). In words, our excluded instruments are <u>two-way</u> interactions between the number of school-age years of exposure to a court-ordered SFR and Head Start spending per four-year-old when the individual was age 4. Because a school district may be a smaller unit of observation than a county, all models include district fixed effects (which subsumes the childhood county fixed effects). The resulting model is as in [6], where $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ are fitted values from first-stage regressions.[4] Note that this is no longer an overidentified model, but is a just identified model.

[6] $$Y_{icb} = \beta_{HS} \cdot HS_{cb}^{age\,4} + \beta_{k12} \cdot \widehat{ppe}_{idb}^{5-17} + \beta_{int} \cdot (\widehat{INT}_{idb}) + \gamma \cdot C_{icb} + \theta_d + \tau_{bdi} + \varepsilon_{idb}.$$

The results from this just identified model are presented in Appendix Table H2. As one can see, for our two main outcomes (years of education and wages), the results are very similar to those of the DiD-2SLS models. Also, formal statistical tests fail to reject that these models are the same. Importantly, (even though the standard errors are considerably larger in the simple model) our main findings are robust to this more parsimonious model. This suggests that (a) our measure of dosage is not biased and, (b) our estimated effects are robust to using approaches that do not have multiple first stages.

No selection or endogenous mobility: Another concern one may have with the estimates is that due to selective migration or neighborhood change, the characteristics of the individuals exposed to different levels of K12 spending or Head Start spending are not the same. We address this possible concern in two ways. First, we demonstrate that the spending changes we exploit are unrelated to observed family and neighborhood characteristics. Specifically, we regress year of educational attainment and the adult wage on several observable characteristics and then take the fitted values from those regressions as our predicted outcomes. To obtain these predicted outcomes, we estimated models that predict educational attainment and adult earnings using parental income, race, mother's and father's education and occupational prestige index, mother's marital status at birth, birth weight, childhood county-level average per-capita expenditures on

[4]Where $\hat{X}_1 = \widehat{ppe}_{idb}^{5-17}$ and $\hat{X}_2 = \widehat{INT}_{idb}$, and $w \in \{1,2\}$,
$\hat{X}_w = \pi_{w2}(SFRExp_{idb}) + \pi_{w4}(SFRExp_{idb}) \cdot HS_{cb}^{age\,4} + \gamma_w C_{idb} + \theta_{wd} + \tau_{wb}$.

Title I, AFDC, Medicaid, food stamps, and UI, respectively, during childhood years. The predicted outcomes from these models are intended to capture an effect-size weighted index of childhood family/community SES factors. We then regress our predicted outcomes on the spending changes (excluding all of these same observable characteristics). If the spending changes are unrelated to those observable characteristics that predict the adult outcome, the estimated coefficients will be zero. Indeed, this is what we find (See Table H3).

Even though our spending changes are unrelated to observed characteristics, one may worry about selection on unobserved characteristics. To rule out the possibility that our results are driven by differences across treated and untreated families, we rely on variation within families and compare the outcomes of siblings who were different ages at Head Start rollout or at the time of a court-ordered SFR, but were raised in the same household. This approach accounts for all observed and unobserved shared family characteristics that influence outcomes. We achieve this by augmenting [6] and [7] to include sibling fixed effects. As one can see in Table H4, while we lose considerable precision, the estimated coefficients for low-income children are very similar to those without sibling fixed effects. This suggests that family selection cannot explain the main pattern of results. These sibling tests also address any potential lingering concerns regarding endogenous mobility driving the results, because individuals in the same family have the same residential address. As an additional check on endogenous mobility, we re-estimated our preferred DiD-2SLS models limiting the analysis sample to those who lived at their (earliest) childhood residence before the enactment of Head Start programs in their respective county. NOTE: This does not exclude movers; we exclude the 3% of our sample for whom the initial address *could have been* the result of endogenous movement. The results are presented in Appendix Table H5. As one would expect, we find nearly identical results as those in the full sample. This indicates that endogenous residential mobility is not a major source of bias in this analysis.

Testing for Sufficient Variation to Identify the Interaction Term

Identification of our parameter of interest is based on the interaction between two policy instruments. Credible identification of our parameter requires that there be exogenous variation in both Head Start spending and K12 spending conditional on the other. This issue is discussed in Buckles, Morrill, Hagerman, Wozniak and Malamud (2013). Intuitively, if the same areas that receive increased K12 spending due to reforms are also those that experienced the largest increases in Head Start spending, then there may be no credible exogenous variation in K12 spending conditional on Head Start spending and *vice versa*. With a very high correlation between the two policy instruments, our model would be underidentifed.

We assess whether this is a problem in two different ways. First, we compute the correlation between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) at the childhood county-birth cohort level. If our policy-induced variation in Head Start spending and K12 spending were based on the same sample of counties, there would be a large positive correlation. In fact, the raw correlation (i.e. with no controls) between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) is only 0.15. To test this formally, we ran our 2SLS model predicting Head Start spending at age four as a function of the SFR-induced changes in K12 spending with all the controls from our main specification. The results are presented in Table H6. In such models, the coefficient is an economically insignificant 0.013 and the *p*-value is larger than 0.1. Taken at face value, the point estimate indicates that an exogenous 10% increase in K12 spending is associated with a mere additional $1.3 per poor four-year-old spent on Head Start. Similarly, we regressed the reform-induced change in K12 spending (the fitted

values from the first-stage regression predicting K12 spending) on Head Start spending at age 4. In such models, the coefficient is less than 0.001 with a *p*-value greater than 0.1. In sum, the two sources of exogenous variation are largely unrelated to one another, such that the interaction between the two is identified.

As a further check that there is sufficient variation to uniquely identify each of our endogenous regressors, we follow Angrist and Pischke (2009). To test for sufficient unique variation in our main models that rely on difference-in-difference variation in Head Start spending and instrument for both K12 spending and the interaction between K12 spending and Head Start spending, we report a series of F-statistics (see Table H7). Looking at predicting K12 spending, the first stage F-statistic for the log of K12 spending (based on predicted district-level dosage times years of SFR exposure in the state) is 22.41 and 23.01 in models without and with Head Start variables, respectively. As such, there is a strong first stage for K12 spending whether Head Start spending is included in the model or not. Looking at predicting head Start spending, the first stage F-statistic for the Head Start spending (based on rollout) is 59.17 and 60.76 in models without and with K12 variables, respectively. As such, there is a strong first stage for Head Start spending whether K12 spending is included in the model or not. Also, as a direct test of the strength of the first stage for the interaction in our main models, the first stage F-statistic for Head Start spending times SFR dosage times SFR exposure is 28.71, conditional on Head Start spending and SFR dosage times SFR exposure. Similarly, the first stage F-statistic for Head Start rollout times SFR dosage times SFR exposure is 42.46, conditional on Head Start rollout and SFR dosage times SFR exposure. That is, the F-statistic on the interaction between the two policy instruments in predicting the interaction between the two spending types is large (conditional on the effect of the individual policy instruments themselves). In sum, all the tests indicate that we have sufficient independent exogenous variation to credibly identify the effects of Head Start spending, the effect of K12 spending, and the effects of the interaction between the two.

Appendix Table H1:

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 (Poor children. Outcomes are measured between ages 20-45)

	Years of Education	Prob(High School Grad)	Ln(Wage)	Prob(poverty)	Prob(Ever incarcerated)
	1	2	3	4	5
$Ln(PPE_d)_{(age\ 5\text{-}17)}$	4.9251**	0.9026*	1.3588**	-0.8803**	-0.6448+
	(2.4230)	(0.5430)	(0.6351)	(0.4256)	(0.4003)
$Ln(PPE_d)_{(age\ 20\text{-}24)}$	-0.8152	0.00044	-0.1450	0.0261	0.06397
	(2.5142)	(0.4131)	(0.2805)	(0.1864)	(0.1711)
Number of person-year observations			55,706	88,124	
Number of Individuals	5,419	5,419	5,613	6,373	4,536

Robust standard errors in parentheses (clustered at 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. 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 $dose_d$).

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, rollout 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,% black, education,% urban, population size,% voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends.

Table H2:

Using only School Finance Reform Exposure as Instruments for K12 Spending (Poor Children):

	1	2	3	4
	Years of Education DiD-2SLS		Ln(Wage), ages 20-50 DiD-2SLS	
Head Start Spending$_{(age\ 4)}$	0.07721***	0.0670***	0.02334***	0.0215***
	(0.01992)	(0.0189)	(0.004503)	(0.004523)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	4.0399**	9.4546**	2.0561***	2.7146***
	(1.6751)	(4.4808)	(0.4348)	(0.8493)
Head Start Spending$_{(age\ 4)}$*Instrumented ln(PPE) $_{(age\ 5\text{-}17)}$	0.6460***	0.6879***	0.1698**	0.1660**
	(0.2354)	(0.2485)	(0.06985)	(0.0705)
SFR Exposure and dosage Instruments for K12 Spending?	YES	NO	YES	NO
Only SFR Exposure Years as Instruments for K12 Spending?	--	YES	--	YES
Number of Person-year Observations	--	--	55,706	55,706
Number of Children	5,419	5,419	5,613	5,613
Number of Childhood Families	2,133	2,133	2,202	2,202
Number of School Districts	749	749	761	761
Number of Childhood Counties	600	600	610	610

Robust standard errors in parentheses (clustered at childhood state level)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes as predictors the school-age years of exposure to school finance reform. Results in columns (2) & (4) DO NOT include school finance reform "dosage" intensity terms as instruments (i.e., without the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories), while columns (1) & (3) do include SFR "dosage" as in preferred 2SLS-DiD presented in Table 1a. There exists a significant first-stage.

Table H3.

Examining Exogeneity of Head Start and K-12 Spending (Poor Children)

	Predicted Years of Education, based on Childhood Family & County SES		Predicted Ln(Wages) at age 30, based on Childhood Family & County SES	
	1	2	3	4
	School District FE & Race*Birth Yr FE	Partial Set of Controls	School District FE & Race*Birth Yr FE	Partial Set of Controls
Head Start Spending$_{(age\ 4)}$/1000	-0.0044313	-0.0037102	0.0002552	-0.000113
	(0.0046289)	(0.0048768)	(0.0007137)	(0.0007671)
Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$	0.7713586	0.7905367	0.0432606	0.0438955
	(0.8242292)	(0.8474821)	(0.1618254)	(0.1904868)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes PSID individuals born 1950-1976, followed into adulthood through 2015. We estimated models that predict educational attainment & adult earnings 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 birth, birth weight, childhood county-level average per-capita expenditures on Title I, AFDC, Medicaid, food stamps, & UI, respectively, during childhood years)—this is intended to capture an effect-size weighted index of childhood family/community SES factors. We then examined whether individuals' predicted educational attainment, and wages at age 30 based only on childhood family/county characteristics (i.e., the effect-size weighted index of childhood family/community SES factors) is related to county Head Start spending per poor 4-year old, holding constant school district fixed effects and birth year fixed effects. The results presented in this Table show that, holding constant school district fixed effects and birth year fixed effects, identifying variation in Head Start spending increases are NOT significantly related to (changes in) childhood family/community SES characteristics. Head Start spending per poor 4-year old is in thousands of dollars (real 2000 dollars), so that a one-unit change represents a $1,000 change in spending.

Table H4: *Within Family Model: 2SLS-Difference-in-Difference Estimates of Early and K12 education Spending on Adult Outcomes (Poor Children)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Prob(High School Grad)		Years of Completed Education		Ln(Wage), ages 20-50		Annual Incidence of Poverty, age 20-50		Prob(Ever Incarcerated)	
	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV
Head Start Spending$_{(age\ 4)}$	0.02082*	0.08133**	0.1392***	0.2519*	0.01269**	0.0708*	-0.004733+	-0.02100	-0.0206***	-0.0281
	(0.01252)	(0.04115)	(0.03635)	(0.1525)	(0.005874)	(0.0426)	(0.002972)	(0.02398)	(0.0051)	(0.0343)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	1.5534**	1.2292***	6.0377***	4.5989**	0.9450*	1.3371***	-0.5449*	-0.9089***	0.2587	-0.1613
	(0.7661)	(0.4381)	(2.2362)	(2.0073)	(0.5000)	(0.4016)	(0.3273)	(0.3334)	(0.2720)	(0.4227)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.1062*	0.2374**	0.8159***	0.9395**	0.06292	0.07938	-0.06652+	-0.2153***	0.0744	-0.0360
	(0.05974)	(0.1047)	(0.2958)	(0.4759)	(0.07046)	(0.1091)	(0.04744)	(0.06991)	(0.0462)	(0.1047)
Number of Person-year Observations	5419	5419	5419	5419	55706	55706	88124	88124	4536	4536
Number of Families	2133	2133	2133	2133	2202	2202	2301	2301	1727	1727

Robust standard errors in parentheses (clustered at childhood state level)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Sample includes all individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: (non-Instrumented & Instrumented) Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that include: childhood family fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for child-specific family characteristics (parental income, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. The instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only). There exists a significant first-stage.

Table H5:

Early Address Sample:

Difference-in-Difference-2SLS Estimates of Early and K12 education Spending on Adult Outcomes

	Children from Poor Households				
	1	2	3	4	5
	Years of Education	Prob(High School Grad)	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Incarceration)
Head Start Spending$_{(age\ 4)}$/1000	0.06754***	0.01420**	0.02198***	-0.01649***	-0.005557*
	(0.01584)	(0.006972)	(0.004307)	(0.004463)	(0.003232)
Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$	4.9596***	1.0533**	2.3668***	-0.8955**	-0.7171**
	(1.8986)	(0.4259)	(0.5023)	(0.3964)	(0.3188)
Interaction	0.6981***	0.1316**	0.1749***	-0.1036***	-0.05540*
	(0.2470)	(0.06662)	(0.06136)	(0.03451)	(0.03313)
Number of Person-year Observations	--		53,970	84,326	--
Number of Children	5,071	5,071	5,280	5,971	4,408

Robust standard errors in parentheses (clustered at childhood state level)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, who have been followed into adulthood, and for whom earliest available address predates Head Start rollout and school finance reform.

Models: Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories.

Table H6:

The Relationship between School Finance Reform-Induced Changes in Per-Pupil K12 Spending and Head Start Spending

(**Children from Poor Households**)

	Dependent variable:	
	(1)	(2)
	County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s)	(SFR) Instrumented Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$
County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s)	--	0.0001633
		(0.0003878)
(SFR) Instrumented Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$	0.0133935	--
	(2.853857)	
Number of Children	5,419	5,419

Robust standard errors in parentheses (clustered at childhood state level)

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

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

At the childhood county-birth cohort level, the correlation between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) is 0.15; and controlling for birth year, there is no significant relationship.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, Medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size,% voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage.

Table H7:

F-Statistics on Excluded Instrument in Different Models (Poor children only)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$		Head Start Spending*Ln(K12 Spending)		County Head Start Spending per Poor 4-year old$_{(age\ 4)}$	
Model	without Head Start Spending	with Head Start Spending and SFR dosage*SFR exposure*Head Start spending	With Head Start Spending and SFR dosage*SFR exposure	With Head Start Exposure and SFR dosage*SFR exposure	without SFR dosage*SFR exposure	with SFR dosage*SFR exposure and SFR dosage*SFR exposure*Head Start Exposure
Excluded Instruments	SFR dosage*SFR exposure	SFR dosage*SFR exposure	Head Start Spending*SFR dosage*SFR exposure	Head Start Exposure*SFR dosage*SFR exposure	Head Start Exposure	Head Start Exposure
F-Statistic on excluded instruments	22.41	23.01	28.71	42.46	59.17	60.76

Robust standard errors clustered at childhood state level.

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around its mean to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, food stamps, Medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). The first-stage model of K12 spending include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. The instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only); and in column (6) this instrument is interacted with school-age years of exposure to school finance reform*dosage (reform-induced change in school spending based on the timing and type of court-ordered reform).

Appendix Table H8. *Placebo Tests: Effects of Head Start Spending by Child Age on Educational Attainment, Low-Income Children*

	Completed Years of Education									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age 4: Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s)	0.07640***	0.07742***	0.1077**	0.07392***	0.07000***	0.08065***	0.07577***	0.07830***	0.08185***	0.07734***
	(0.01967)	(0.01942)	(0.04214)	(0.02337)	(0.01599)	(0.02726)	(0.01961)	(0.02009)	(0.01997)	(0.01981)
Age1: Head Start Spending per Poor 4-year old$_{(age\ 1)}$ (in 000s)		0.01329								
		(0.03527)								
Age2: Head Start Spending per Poor 4-year old$_{(age\ 2)}$ (in 000s)			-0.07522							
			(0.05727)							
Age3 Head Start Spending per Poor 4-year old$_{(age\ 3)}$ (in 000s)				0.01079						
				(0.02878)						
Age5: Head Start Spending per Poor 4-year old$_{(age\ 5)}$ (in 000s)					0.02417					
					(0.03274)					
Age6: Head Start Spending per Poor 4-year old$_{(age\ 6)}$ (in 000s)						-0.01763				
						(0.05074)				
Age7: Head Start Spending per Poor 4-year old$_{(age\ 7)}$ (in 000s)							0.005178			
							(0.02362)			
Age8: Head Start Spending per Poor 4-year old$_{(age\ 8)}$ (in 000s)								-0.009519		
								(0.02046)		
Age9: Head Start Spending per Poor 4-year old$_{(age\ 9)}$ (in 000s)									-0.05718**	
									(0.02868)	
Age10: Head Start Spending per Poor 4-year old$_{(age\ 10)}$ (in										-0.008806
										(0.01934)
Number of Children	5,378	5,378	5,378	5,378	5,378	5,378	5,378	5,378	5,378	5,378
Number of School Districts	761	761	761	761	761	761	761	761	761	761
Number of Childhood Counties	577	577	577	577	577	577	577	577	577	577

Robust standard errors in parentheses (clustered at childhood

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

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is measured in 000s. These results are also presented in Figures 7a-7f across all outcomes. Results are based on Difference-in-Difference models that include same full set of controls (as in Tables 1-2): parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, Medicaid, AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size,% voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender).

Appendix Table H9. *Placebo Tests: Effects of Head Start Access by Child Age on Educational Attainment, Low-Income*

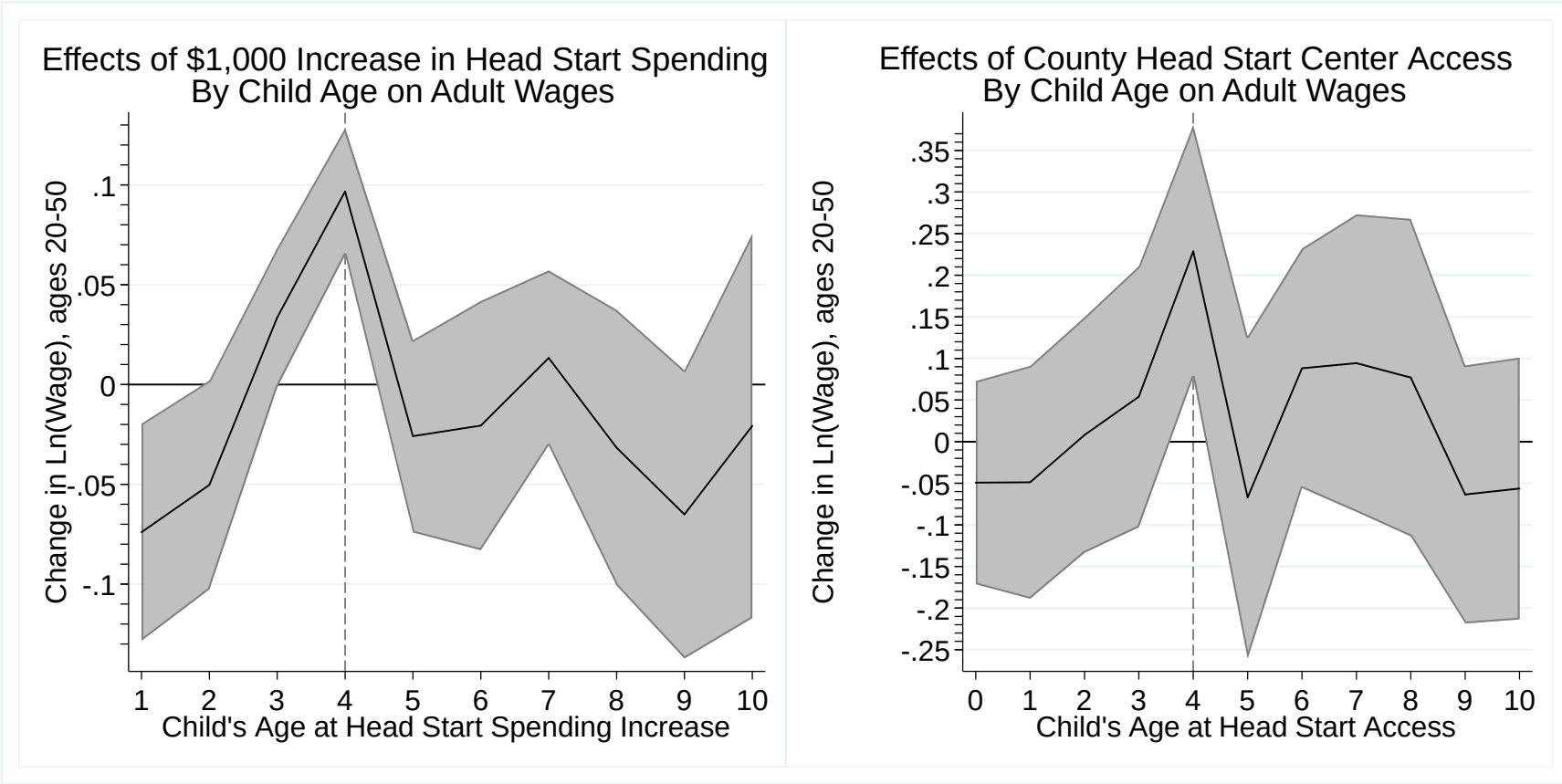
	Completed Years of Education									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age 4: Head Start Center	0.8889**	0.9009**	0.8228*	1.1938**	1.2516**	1.0180**	0.7025*	0.7774**	0.7791**	0.7402**
	(0.3710)	(0.4065)	(0.4972)	(0.5271)	(0.5001)	(0.4340)	(0.4025)	(0.3552)	(0.3599)	(0.3363)
Age1: Head Start Center		-0.02754								
		(0.2867)								
Age2: Head Start Center			0.08828							
			(0.4110)							
Age3 Head Start Center				-0.4204						
				(0.3922)						
Age5: Head Start Center					-0.4697					
					(0.4667)					
Age6: Head Start Center						-0.2205				
						(0.4303)				
Age7: Head Start Center							0.3626			
							(0.3297)			
Age8: Head Start Center								0.2284		
								(0.3684)		
Age9: Head Start Center									0.2445	
									(0.4334)	
Age10: Head Start Center										0.4148
										(0.4142)
Number of Children	5,378	5,378	5,378	5,378	5,378	5,378	5,378	5,378	5,378	5,378
Number of School	761	761	761	761	761	761	761	761	761	761
Number of Childhood	577	577	577	577	577	577	577	577	577	577

Robust standard errors in parentheses (clustered at childhood county level)
*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is measured in 000s. These results are also presented in Figures 7a-7f across all outcomes. Results are based on Difference-in-Difference models that include same full set of controls (as in Tables 1-2): parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, Medicaid, AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size,% voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender).

Figure H1: *Effect of Head Start Spending and Rollout on Adult Wages by Age*



These figures present the marginal effects of Head Start spending in an individual's childhood county at different ages, conditional on the level of Head Start spending in the childhood county at age 4 (when such spending should have an effect). The sample is poor children only. Models include the set full set of controls as in Tables 2 and 3. The coefficients on the non-eligible years 1 through 3 and 5 through 10, are all conditional on spending at age 4. The coefficient for spending at age 4 is based on a model with no other ages included.

Appendix I: *Testing for Improvement in Parent Quality due to Head Start*

Table I1:
Test for any Spillover Effects of Head Start Spending on Older Sibling not Exposed at age 4

	Children from Low-Income Families				
	1	2	3	4	5
	Prob(High School Grad)	Years of Education	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Ever Incarcerated)
Younger Sibling's County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s)	-0.0233	0.0118	0.0024	-0.0033	-0.0021
	(0.0206)	(0.0690)	(0.0269)	(0.0172)	(0.0053)

Robust standard errors in parentheses (clustered at childhood state level)
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes only older siblings not exposed to Head Start (i.e., who turned age 4 before the program's rollout), but whose younger sibling(s) had a Head Start center in the county when they were age 4, and whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.
Models: Head Start spending per poor 4-year old in the county (measured in 000s). Results are based on models that include same set of controls as Tables 1-2: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, food stamps (average during age 0-4), Medicaid (average during age 0-4), AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic).

Appendix J

Table J1: *Poor vs Non-Poor Children (DiD-2SLS Models)*

	(1)	(2)	(3)	(4)	(5)	(6)
	Years of Education			Ln(Wage), ages 20-50		
	Poor	Non-Poor	Difference	Poor	Non-Poor	Difference
Head Start Spending$_{(age\ 4)}$	0.07721***	0.008866	0.0683***	0.02334***	0.006901	0.0164**
	(0.01992)	(0.01635)	(0.0258)	(0.004503)	(0.005408)	(0.0070)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	4.0399**	2.4192**	1.6207	2.0561***	0.7351**	1.3210**
	(1.6751)	(1.1645)	(2.0401)	(0.4348)	(0.3035)	(0.5302)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.6460***	0.02972	0.6163**	0.1698**	0.02577	0.1440*
	(0.2354)	(0.1937)	(0.3048)	(0.06985)	(0.03090)	(0.0764)
Number of Person-year Observations	--	--		55,706	90,771	
Number of Children	5,419	7,983		5,613	8,195	
Number of Childhood Families	2,133	3,530		2,202	3,593	
Number of School Districts	749	1,156		761	1,169	
Number of Childhood Counties	600	891		610	908	

Robust standard errors in parentheses (clustered at childhood state level)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 who have been followed into adulthood (218,594 person-year observations; 15,232 individuals; 4,990 childhood families; 1,427 school districts; 1,120 childhood counties).

Models: Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on DiD-2SLS models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage.

Appendix K:

Table K1: *K12 Spending Effects by Age at First School Finance Reform*

Dependent Variable:	Years of Education	Ln(Wage), ages 20-50	Years of Education	Ln(Wage), ages 20-50
	Exposed to SFR 1-10 years: (after 7)		Exposed to SFR 11+ years: (before 7)	
(SFR) Instrumented $Ln(PPE)_{(age\ 5\text{-}17)}$	4.3942*	2.1295**	2.1434	0.8305**
	(2.3134)	(0.9008)	(2.0695)	(0.3604)
Head Start $Spending_{(age\ 4)}$*$ln(PPE)_{(age\ 5\text{-}17)}$	0.1900	0.3313*	0.6331**	0.1338**
	(0.3610)	(0.1774)	(0.2591)	(0.0630)
	Exposed to SFR 1-9 years: (after 8)		Exposed to SFR 10+ years: (before 8)	
(SFR) Instrumented $Ln(PPE)_{(age\ 5\text{-}17)}$	7.0847**	2.6364***	4.5313***	1.2779***
	(3.2518)	(0.9405)	(1.1263)	(0.3151)
Head Start $Spending_{(age\ 4)}$*$ln(PPE)_{(age\ 5\text{-}17)}$	0.3217	0.3512	0.6683**	0.1423**
	(0.4321)	(0.2376)	(0.2745)	(0.0662)
	Exposed to SFR 1-8 years: (after 9)		Exposed to SFR 9+ years: (before 9)	
(SFR) Instrumented $Ln(PPE)_{(age\ 5\text{-}17)}$	6.3156+	1.8053**	3.8329***	1.3591***
	(4.8084)	(0.7514)	(1.1443)	(0.4269)
Head Start $Spending_{(age\ 4)}$*$ln(PPE)_{(age\ 5\text{-}17)}$	0.0590	0.1491	0.6054***	0.1553**
	(0.8043)	(0.1113)	(0.2230)	(0.0665)

*** p<0.01, ** p<0.05, * p<0.1 Robust standard errors in parentheses (clustered at childhood state level)

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Each panel is based on a single regression in which all the K12 instruments and K12 spending variables are interacted with indicator variables connoting whether an SFR occurred in their childhood state before a particular age or after a particular age (top panel is 7, middle panel is 8, and the bottom panel I 9 years old). Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, and both the county per-capita Medicaid and food stamps spending variables are also included and centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, food stamps (average during age 0-4), Medicaid (average during age 0-4), 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 cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories; and each of these variables interacted with an indicator for whether individual was exposed to an SFR before or after 7 (top panel), 8 (middle panel), or 9 (bottom panel) years old. There exists a significant first-stage for all variables in all models.

Table K2: *First Stage Estimates*

	1	2	3	4	5
Endogenous Variable (Dependent Variable of First Stage)	Ln(PPE)$_{(age\ 5\text{-}17)}$		Head Start Spending*Ln(PPE)		Head Start Spending
Model	DID-2SLS	2SLS-2SLS	DID-2SLS	2SLS-2SLS	2SLS-2SLS
Head Start Spending Instruments					
Head Start Exposure$_{(age4)}$		-0.0487		0.3228+	3.0333***
		(0.0516)		(0.2088)	(0.3891)
K12 Spending Instruments					
# of School-age years of SFR exposure	0.0091*	0.0291***	0.0387+	-0.0860+	0.0830
	(0.0048)	(0.0086)	(0.0261)	(0.0629)	(0.1217)
(School-age years of SFR exposure)*(Dosage quartile2)	0.0049	0.0041	-0.0165	-0.1242**	-0.2763***
	(0.0047)	(0.0076)	(0.0273)	(0.0537)	(0.0820)
(School-age years of SFR exposure)*(Dosage quartile3)	0.0053+	0.0032	-0.0781***	-0.0567***	-0.0080
	(0.0035)	(0.0054)	(0.0153)	(0.0204)	(0.0831)
(School-age years of SFR exposure)*(Dosage top quartile)	0.0110***	-0.0454***	-0.0819***	0.1548***	-0.0592
	(0.0033)	(0.0084)	(0.0261)	(0.0314)	(0.1476)
Instruments for Interaction (DID-2SLS models)					
# of School-age years of SFR exposure*Head Start spending$_{(age4)}$	-0.0009		0.0155***		
	(0.0007)		(0.0034)		
(School-age years of SFR exposure)*(Dosage quartile2)*Head Start	-0.0068***		0.0294***		
	(0.0015)		(0.0098)		
(School-age years of SFR exposure)*(Dosage quartile3)*Head Start	0.0015		-0.0136**		
	(0.0014)		(0.0057)		
(School-age years of SFR exposure)*(Dosage top quartile)*Head Start	0.0025		-0.0160+		
	(0.0017)		(0.0109)		
Instruments for Interaction (2SLS-2SLS models)					
# of School-age years of SFR exposure*Head Start Exposure$_{(age4)}$		-0.0249***		0.1390***	-0.0156
		(0.0048)		(0.0284)	(0.0399)
(School-age years of SFR exposure)*(Dosage top quartile)*Head Start		0.0614***		-0.2423***	0.0429
		(0.0060)		(0.0205)	(0.0834)
Number of Children	5,419	5,419	5,419	5,419	5,419

Robust standard errors in parentheses (clustered at childhood state level)
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born

Models: The set of controls include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, roll-out of "War on Poverty" & related safety-net programs (community health centers, food stamps, Medicaid, AFDC, UI, Title-I (average during childhood yrs)), 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 cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). The first-stage model of K12 spending include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. The instrument used for Head Start spending per poor 4-year old is an indictor for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only); and this instrument is interacted with school-age years of exposure to school finance reform*dosage (reform-induced change in school spending based on the timing and type of court-ordered reform). There exists a strong first-stage in all models.

Table K3: *2SLS-DiD Estimates With and Without Controls for Other Policies: Poor Children*

	1	2	3	4
	Years of Education		Ln(Wage), ages 20-50	
	Parsimonious	Full Set of Controls	Parsimonious	Full Set of Controls
Head Start Spending$_{(age\ 4)}$	0.0751***	0.07721***	0.0226***	0.02334***
	(0.0221)	(0.01992)	(0.0042)	(0.004503)
Ln(PPE)$_{(age\ 5\text{-}17)}$	3.4683**	4.0399**	1.6507***	2.0561***
	(1.6641)	(1.6751)	(0.5387)	(0.4348)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.6227***	0.6460***	0.1787***	0.1698**
	(0.2208)	(0.2354)	(0.0656)	(0.06985)
Number of Person-year Observations	--	--	55,706	55,706
Number of Children	5,419	5,419	5,613	5,613

Robust standard errors in parentheses (clustered at childhood state level)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV-DiD models. Columns (1), (3), (5) include a more parsimonious set of controls that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends, and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), age (cubic). Columns (2), (4), (6) use complete set of controls that in addition to the aforementioned include: controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood. The first-stage model includes as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories.

Table K4: *OLS Estimates of Interactive Effects of Head Start and K12 Spending (Poor Children)*

	1	2	3	4	5
	Years of Education	Prob(High School Grad)	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Ever Incarcerated)
Head Start Spending$_{(age\ 4)}$	0.07965***	0.02570***	0.01327***	-0.01116***	-0.006562*
	(0.01271)	(0.006439)	(0.004868)	(0.002061)	(0.003698)
Ln(PPE)$_{(age\ 5\text{-}17)}$	0.6542	0.2283	0.2537**	-0.07677	-0.05556
	(0.5300)	(0.1823)	(0.1228)	(0.06654)	(0.07668)
Head Start Center$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.2473**	0.04139	0.04435**	-0.02757**	-0.004091
	(0.1210)	(0.02703)	(0.02135)	(0.01063)	(0.01583)
Number of Person-year Observations	--	--	55,706	88,124	--
Number of Children	5,419	5,419	5,613	6,373	4,536

Robust standard errors in parentheses

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, who have been followed into adulthood.

Models: Head Start Spending per poor 4-year old at age 4 in the county and ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean. Results are based on OLS models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic).

Table K5: *OLS Estimates of Interactive Effects of Head Start and K12 Spending (Non-Poor Children)*

	1	2	3	4	5
	Years of Education	Prob(High School Grad)	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Ever Incarcerated)
Head Start Spending$_{(age\ 4)}$	0.01671	0.00094	0.006004	-0.00088	-0.000725
	(0.01845)	(0.003676)	(0.004257)	(0.001115)	(0.001451)
Ln(PPE)$_{(age\ 5\text{-}17)}$	-0.2178	0.01140	0.02731	-0.009794	0.05600
	(0.2585)	(0.06029)	(0.08434)	(0.01899)	(0.03661)
Head Start Center$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	-0.07470	-0.002712	0.01167	-0.001164	0.007585
	(0.05749)	(0.01116)	(0.01642)	(0.003119)	(0.004769)
Number of Person-year Observations	--	--	90,771	130,470	--
Number of Children	7,983	7,983	8,195	8,859	5,140

Robust standard errors in parentheses

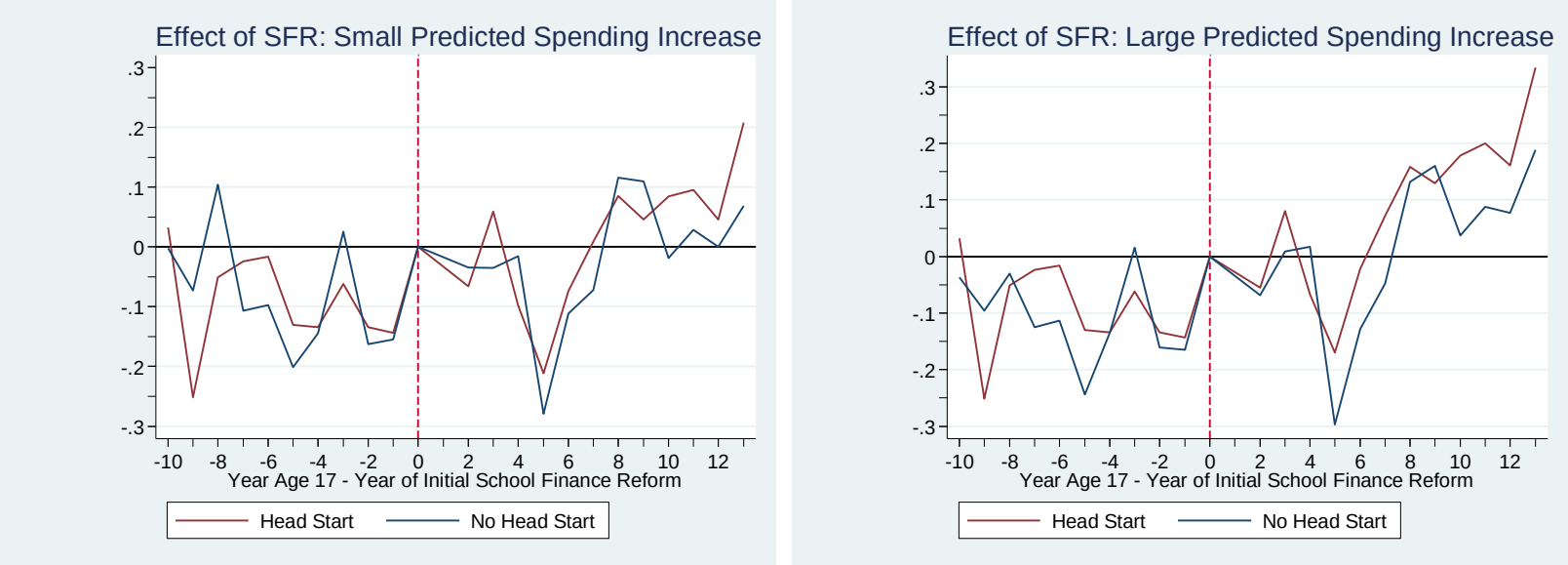
*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were NOT in the bottom quartile of the income distribution, who have been followed into adulthood.

Models: Head Start Spending per poor 4-year old at age 4 in the county and ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean. Results are based on OLS models that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic).

Appendix L

Figure L1: *Interaction Event Study for Wages*



Models: The event study figures use school district's predicted reform-induced change in spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories--right panel shows estimated effects for districts with a predicted reform-induced K12 spending increase ($\widehat{dose}_d > 0$) whereas the left panel shows the corresponding effects for districts with low predicted reform-induced K12 spending increases or a decrease $\widehat{dose}_d \leq 0$. Roughly two-thirds of districts in reform states had predicted spending increases. These estimated effects are presented both for children whose county had no Head Start center at age 4 (blue line), and those who were exposed to any county Head Start spending at age 4 (red line), to highlight the role of dynamic complementarity. Transitory fluctuations in adult wages may introduce more noise to these event study figures than comparable ones for educational attainment (as presented in Figure 5). The event study models include: school district fixed effects, race-specific year of birth fixed effects, age (cubic), controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender) and same set of other controls as main models.

Appendix M

Estimating Head Start Participation Rates

The ratio of enrolled students to the income-eligible age-eligible population in a given year *is not* the same as a specific cohort's participation rate by kindergarten entry.

To relate these enrollments to participation rates at the individual child level, for each kindergarten entry cohort we computed the cumulative likelihood across all age-eligible years that an income-eligible child would enroll in Head Start.

To illustrate this point, suppose for simplicity that Head Start had only the summer program. For example, the annual enrollment rate in summer programs was about 22 percent between 1965 and 1967. The cohort of income-eligible children entering kindergarten in 1965 could only have enrolled at age 5 and would have a 22 percent participation rate. However, the cohort of income-eligible children that entered kindergarten in 1966 could have enrolled at age 4 or 5, so (assuming that participants enroll for one year and not multiple years) their cohort's participation rate by kindergarten entry would be 44 percent (i.e., the sum of the likelihood of participation during ages 4 or 5: 22 + 22). All subsequent cohorts could have enrolled at ages 3, 4, or 5, so that post-1966 cohorts' participation rate by kindergarten entry (across all age-eligible years) is the running total annual summer enrollment ratio for the three years preceding kindergarten entry. Similarly, assuming that participants enroll for one year and not multiple years, a specific cohort's full-year participation rate by kindergarten entry is the running total annual full-year enrollment ratio for the two years preceding kindergarten entry.

To avoid double-counting individuals who enrolled in both the summer program and the full-year programs, we assume that 40 percent of full-year enrollees were previously in a summer program.

Appendix N

Details of the back of the envelope calculations:

It is helpful to define some parameters. The proportion of poor children in a county is *p*. The average per-student cost of rolling out the average Head Start center is the cost of increasing Head Start spending per poor-4-year old by $4,320. The average cost of this increase is simply 4320**p*. The marginal effect of rolling out the average Head Start center for a county (π_{HS}), is a poverty-weighted average of the effect of a $4,320 increase in Head Start spending on low-income children ($\delta_{HS,\ poor}$), and that for non-poor children ($\delta_{HS,non}$). Because Head Start has no effect on non-poor children, this simplifies to [8] below.

[8] $$\pi_{HS} = p\delta_{HS,poor}.$$

To equate the marginal effects of spending on Head Start to that of spending on the K12 system, we need to define the change in K12 spending that would lead to the same expenditure as an increase of $4,320 in Head Start spending per poor-4-year old. During our sample period, K12 spending was roughly $4,000 per student per year *on average*. Assuming a 7% interest rate, spending $4,000 for 12 years is equivalent to $34,000 in present value terms. Thus, an equivalent expenditure at the *student* level would be a 4320*p*/34000=(*p**12.7)% increase in K12 spending. We define $\delta_{K12,poor}$ and $\delta_{K12,non}$ as the effect of increasing K12 spending by one% on poor and non-poor children, respectively. The marginal effect of the equivalent increase in K12 spending on the average child in the county is therefore

[9] $$\pi_{K12} = (p\delta_{K12,poor} + (1-p)\delta_{K12,non})(12.7p)$$

The ratio shown in [10] between these two equations π_{HS}/π_{K12} is the relative effectiveness of rolling out Head Start (from having no center) and spending the same amount across all children from that same cohort in the county in the K12 system.

[10] $$\frac{\pi_{HS}}{\pi_{K12}} = \frac{p\delta_{HS,poor}}{(p\delta_{K12,poor}+(1-p)\delta_{K12,non})(12.7\text{p})} = \frac{\delta_{HS,poor}/(12.7)}{p(\delta_{K12,poor}-\delta_{K12,non})+\delta_{K12,non}}$$

The relative marginal effect of Head Start rollout and the equivalent spending in the K12 system is a function of the poverty rate *p* as long as $\delta_{K12,poor} \neq \delta_{K12,non}$. Specifically, if $\delta_{K12,poor} > \delta_{K12,non}$, then this ratio is falling in *p*, and if $\delta_{K12,poor} < \delta_{K12,non}$ this ratio is increasing in *p*. Intuitively, if non-poor children are more responsive than poor children to increases in K12 spending (i.e. $\delta_{K12,poor} < \delta_{K12,non}$), then the marginal benefit of increased K12 spending declines with the poverty rate so that the *relative* effectiveness of Head Start spending increases with the poverty rate. The converse is also true.

Appendix O

Additional Contextual Details on Head Start

- Children who are 4 years old and whose family income is below the federal poverty guidelines (or is on public assistance programs AFDC or SSI) are eligible for the program. Beginning in 1972 (as part of the Economic Opportunity Act Amendment) at least 10% of children per center must have a disability (irrespective of the family income of these children). In 1969, a provision was added allowing children from families above the poverty level to receive Head Start services for a fee. A fee schedule for non-poor participants in Head Start was required; fees were prohibited for families below the poverty line. The eligibility criteria was mostly unchanged during the period of the program we analyze (Source: 45 CFR (Code Federal Regulations), Parts 1301 to 1311, Early Childhood Learning and Knowledge Center: http://eclkc.ohs.acf.hhs.gov/hslc; www.eric.ed.gov; Zigler and Valentine, 1979).

- An OEO report of 1967 documents Head Start accomplishments in the first two years on child health that include 98,000 eye defects treated; 900,000 cases of dental problems addressed (5 cavities per child); 740,000 without polio vaccinations received vaccines; and 1,000,000 were given measles vaccinations.