

# Appendix

## A state-level perspective on school spending and educational outcomes

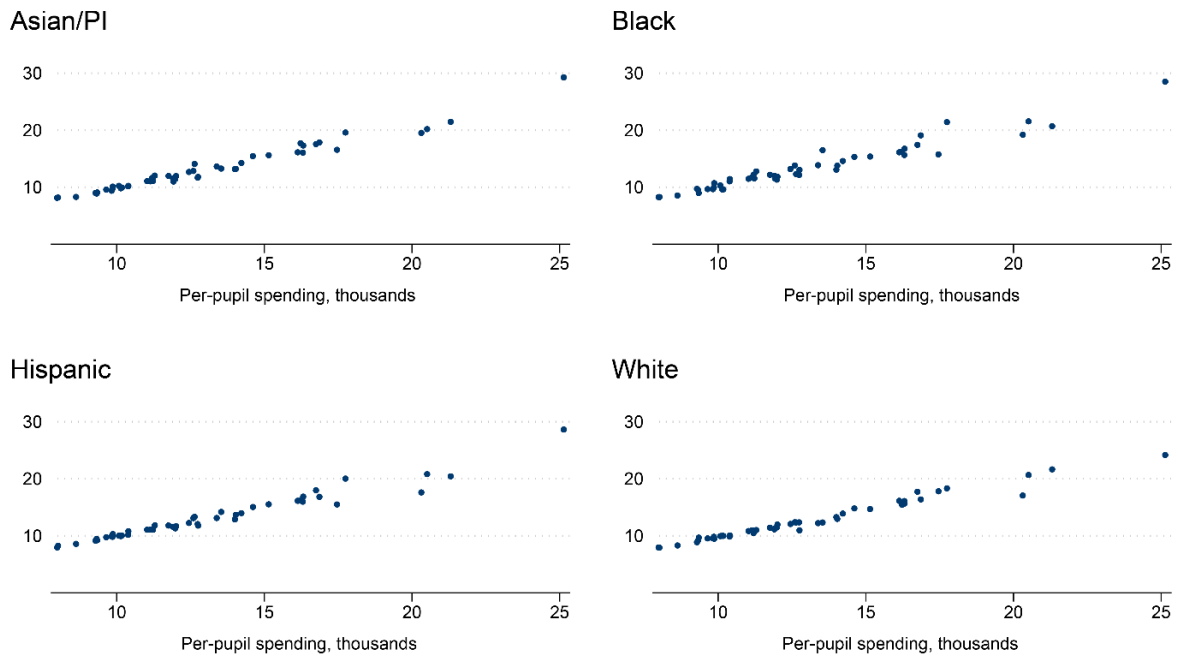
Sarah Reber and Gabriela Goodman  
The Brookings Institution, 2025

**Appendix Table 1: 2018-2019 per-pupil spending predictors**

	(1)	(2)	(3)	(4)	(5)
Per Capita Income (thousands)	354** (43)				388** (71)
Poverty Rate (0-100)		-600*** (135)			592*** (179)
Percent White or Asian (0-100)			34 (28)		100*** (20)
Trump % Two-Party Vote (0-100)				-192*** (49)	-135*** (38)
Constant	-5379** (2117)	20796*** (1956)	11018*** (1845)	23258*** (2681)	-13,669** (6696)
Observations	49	49	49	49	49
R-squared	0.633	0.203	0.023	0.326	0.788
Adj. R-squared	0.625	0.186	0.002	0.311	0.769

*Standard errors in parentheses*  
\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

# Appendix Figure 1. Average total per-pupil spending versus average spending by student race/ethnicity



**Source:** Authors' calculations based on Census Bureau and National Center for Education Statistics (NCES) data.

**Notes:** Per-pupil spending is for the 2018-19 school year. Total per-pupil spending is the weighted average of district-level per-pupil spending, weighted by total enrollment. Race/ethnicity specific per-pupil enrollment is district-level per-pupil spending, weighted by enrollment for each race/ethnicity.

## Data Appendix

### School spending

In this analysis, we use current expenditure per pupil to measure school spending. This excludes debt service and capital expenditure, which can be more volatile year-to-year. Per-pupil revenue, total expenditure, and current expenditure are highly correlated within year at the state level (over 0.98), so the choice of resource measure does not affect the findings. State-level total current expenditure comes from the Annual Survey of Local Governments/F33 and is available starting in 1992; we use the version published by the Census Bureau. Pupil counts are based on fall membership (enrollment) collected as part of the Common Core of Data (CCD) and released together with the F33.

To explore whether how spending is allocated across categories varies with per-pupil spending, we use spending breakdowns also reported in COG/F33. All spending and income measures are adjusted to the 2018 U.S. price level using the CPI-U.

### Test scores

The National Assessment of Education Progress (NAEP), also known as the “Nation’s Report Card” generally administers standardized tests in reading and math to fourth and eighth graders every two years. Tests in other subjects and twelfth grade were also sometimes administered, but not consistently, so we do not consider them here. The sampling procedure for state-level estimates is designed to produce representative samples of public-school students in each state. State-level average test scores are generally available separately by participation in the National School Lunch Program (NSLP) and race, though data are missing when sample sizes are too small to provide reliable estimates. The state NAEP data for fourth and eighth grade reading and math begin in 1990, but not all states participated before 2003 so we drop the earlier years from the analysis (we label these according to the fall of the school year, so we refer to the spring 2003 test as 2002, corresponding to the 2002-03 school year).

NAEP scores for fourth and eighth grade reading and math range from 0 to 500, but standardized test score scales are inherently arbitrary; 10 points on one test might correspond to 50 on another. To facilitate interpretation of effect sizes and comparisons across studies using different test scores as outcomes, most studies (and the JM meta-analysis) standardize scores so that one unit represents a movement of one standard deviation in the student-level distribution of scores. To standardize the NAEP scores, for each grade-year-subject-state-subgroup mean, we subtract the grade-year-subject mean and divide by the grade-year-subject student-level standard deviation.<sup>1</sup>

NAEP administers a standard test to a representative sample, facilitating comparison over time and across states, but the estimates of average test scores can be noisy, especially for smaller states and subgroups. For the last couple decades, federal law has required states to administer standardized tests to most students in certain grades and report the results publicly for some subgroups. In addition to the NAEP data, we use the state-level assessment data from the Stanford Education Data Archive (SEDA), which are equilibrated for comparison of tests scores across states, subjects, and grade-levels. SEDA includes scores for third grade through eighth grade in both math and reading by subgroup. Because these data are based on the universe of students taking the state accountability test and cover more grades, they offer more statistically precise estimates of average achievement. We use the SEDA 5.0 cohort standardized scale data that is disaggregated at the state, subject, and grade level and includes data for 2008 through 2018 (Fahle et al. 2023).

For both NAEP and SEDA, scores are sometimes missing for certain subgroups in certain states. To ensure we have the largest possible sample for the analysis of 2018 data, impute NAEP scores to the

---

<sup>1</sup> As described by Bond and Lang (2013), there are limitations to scaling decisions such as standardization. Because the variance of test scores differs by grade-year-subject, the standardized scales aren’t strictly comparable across grade-year-subject categories.

SEDA data for New Mexico, New York, and Vermont. Then within each dataset (NAEP and SEDA), we fill in missing values by linearly interpolating at the state-grade-subject-subgroup level if the gap between non-missing observations is two years or less. We then average together all available test scores (including those interpolated) for a given state-subgroup-year combination. For instance, the 2018 North Carolina NAEP score is an average of the NAEP fourth grade reading, fourth grade math, eighth grade reading, and eighth grade math average standardized scores in North Carolina in 2018. If the fourth grade math score was missing (and not interpolated as described above), then we would construct the average using the three other score types. This means that the composition of our aggregate average SEDA and NAEP may differ across states and time. If we require a consistent set of grades and subjects for each state-year observation, we will lose some states and years from the analysis. We have explored these trade-offs in supplementary analyses, and our findings are broadly robust to a range of choices about how we construct the data.

## Educational attainment

Measures of educational attainment at the state level are surprisingly difficult to come by. Causal studies tend to use either constructed cohort graduation rates or constructed dropout rates. Measuring high school graduation is difficult because it requires tracking students over time. States are federally mandated to report their Adjusted Cohort Graduation Rate (ACGR), the percentage of students who graduate in four years out of their original cohort of first-time ninth graders (with adjustments made for students who transfer in or out of the cohort). While ACGR is the official graduation rate measure, it is not clear states are using standard procedures, and some manipulation of the data has been uncovered (Harris et al. 2020). We use a five-year average of these data (2014 through 2018), dropping Alabama because of an NCES data quality flag.

We also use two “status” measures of high school dropout based on the 2015 through 2019 American Community Survey (ACS). First, we calculate the percent of 17-year-olds who are not in school and do not have a high school diploma or GED. Most 17-year-olds are still living with their parents and have typically not moved away to college yet, so this is a reasonable proxy for a state’s high school dropout rate for this age group. In theory, the fact that most 17-year-olds still live with their parents makes it possible to calculate this measure separately by parental characteristics. However, even when averaging over five years or more, the sample sizes are too small in many states to produce reliable estimates for the overall population, much less subgroups. To increase the sample size, we also construct a status dropout rate for 17- to 25-year-olds.<sup>2</sup> This increases the sample size, but older students are more likely to have moved from where they attended high school (to go to college, for example); it is difficult to assess the degree of bias this might introduce.

We explore the relationship between spending and each of these measures for the whole population in recent years, but because of the limitations described above, we focus on test scores when examining trends over time and differences across subgroups.

## Staffing and salaries

We use staffing data from the Common Core of Data (CCD). CCD collects and reports annual staff counts by professional category for each state. We average together 2018 and 2019 data in our analysis to create a more reliable estimate.<sup>3</sup> We exclude Ohio from staffing analysis because it seemed to be an implausibly large outlier in the “other staff” category. Our state average teacher salary measure is from the National Education Association and accounts for primary and secondary education teacher salaries.

---

<sup>2</sup> This differs from the age 16 to 24 status dropout rate calculated and reported by NCES. We use the age 17 to 25 range because it excludes younger students who may have not had the chance to drop out yet and accounts for students who may have reengaged with the education system by age 25.

<sup>3</sup> For Nevada, we use the 2019 observation only because the 2018 data are missing for some staffing categories.

## Other state characteristics

In some analyses, we adjust for differences in input prices using the Comparable Wage Index for Teachers (CWIFT) produced by the NCES. CWIFT captures regional variations in the compensation of college graduates who are not pre-K to twelfth grade educators. It is developed using regression analysis based on ACS data. CWIFT is currently available for 2015 to 2022.

We also use state-level data on per-capita income and poverty rates. We derive per-capita income based on personal income and population estimates from the U.S. Census Bureau. We use the overall state-level poverty rates from the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE).

To explore whether the effects of spending on test scores vary by a state's racial composition, we use the state-level percent white and Asian students included in the SEDA data. We use data reported by NCES on the percent of students with an Individualized Education Plan (IEP) and the percent who are English Learners as reported in the Digest of Education Statistics to see how these variables correlate with spending.

## References

- Bond, Timothy N., and Kevin Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K–3: The Fragility of Results." *Review of Economics and Statistics* 95 (5): 1468–79. [https://doi.org/10.1162/REST\\_a\\_00370](https://doi.org/10.1162/REST_a_00370).
- Fahle, Erin M., Sean F. Reardon, Benjamin R. Shear, Andrew D. Ho, Jim Saliba, and Demetra Kalogrides. 2023. "Stanford Education Data Archive Technical Documentation." [https://stacks.stanford.edu/file/druid:dt080zr0625/seda2022\\_documentation\\_2.0.pdf](https://stacks.stanford.edu/file/druid:dt080zr0625/seda2022_documentation_2.0.pdf).
- Harris, Douglas N., Lihan Liu, Nathan Barrett, and Ruoxi Li. 2020. *Is the Rise in High School Graduation Rates Real? High-Stakes School Accountability and Strategic Behavior*. Brookings Institution. <https://www.brookings.edu/wp-content/uploads/2020/02/Is-the-Rise-in-High-School-Graduation-Rates-Real-FINAL.pdf>.