Replication Note: Jackson, Johnson & Persico (2016)

School Finance Reforms and Student Outcomes

Avi Turetsky

Manus AI

(human-AI collaborative working paper)

Additional AI tools: Perplexity Pro, Claude (Anthropic, opus-4)

May 2026 — Working Draft

Human-AI K-12 Evidence Project https://k12evidence.org

Zenodo: https://doi.org/10.5281/zenodo.20109658

Abstract

This note reports the results of an attempted replication of Jackson, Johnson, and Persico (2016), "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *Quarterly Journal of Economics* 131(1): 157–218. Using publicly available data from the American Community Survey (ACS), the Common Core of Data (CCD), and the Stanford Education Data Archive (SEDA), we confirm the direction of the first-stage relationship between finance reforms and per-pupil spending. Full instrumental-variable replication is not feasible with public data: the JJP design requires restricted-use district-level PSID geocodes, and the state-level instrument available in public data is near-collinear with the required fixed effects (first-stage $F = 2.8$ vs. JJP's $F \approx 20$–30). Reduced-form estimates of adult outcomes are near-zero and statistically insignificant, consistent with the weak-instrument problem. An out-of-sample extension using SEDA district-level test score data yields a negative and statistically significant association between post-reform exposure and test scores (-0.117 SD, $p < 0.001$), though this result is confounded by severe sample imbalance and should not be interpreted causally. All code and processed data are archived on

Zenodo.

1 The Original Paper

Jackson, Johnson, and Persico (2016) [1] — hereafter JJP — ask a deceptively simple question: does more money spent on K-12 education actually improve student outcomes? The paper's central contribution is a credible causal answer. Using a sample of 15,353 individuals from the Panel Study of Income Dynamics (PSID), JJP exploit the wave of court-ordered school finance reforms (SFRs) that swept across 28 states between 1971 and 2010. These reforms, triggered by state supreme court rulings that existing funding systems were unconstitutional, forced states to equalize per-pupil spending across wealthy and poor districts. Crucially, the timing of reforms was driven by litigation outcomes rather than by district-level economic trends, providing the exogenous variation needed for causal identification.

The paper's identification strategy is a two-stage instrumental variables (IV) design. The instrument is the predicted change in per-pupil spending for a student's specific childhood school district, based on that district's pre-reform spending level and the timing of the state's court order. Because low-income districts received the largest spending increases (they were the furthest below the new equalized standard), the design effectively compares outcomes for students who grew up in districts that received large reform-induced spending increases against those in districts that received small increases, after controlling for state and birth-year fixed effects.

JJP's core findings are as follows. For the full sample, a 10 percent increase in per-pupil spending throughout the school years (ages 5–17) is associated with 0.31 additional completed years of education, a 7 percent increase in adult wages, and a 3.2 percentage point reduction in the poverty rate in adulthood. These effects are concentrated almost entirely among students who grew up in low-income families: for low-income children specifically, a 10 percent spending increase is associated with a 7 percentage point increase in high school graduation rates, 9.6 percent higher earnings, and a 6.1 percentage point reduction in poverty. The effects are large in magnitude and precisely estimated, with IV first-stage F-statistics in the range of 20–30.

2 What We Attempted to Replicate, and What Constrained Us

2.1 The Binding Data Constraint

The JJP design requires linking each PSID respondent to their specific childhood school district. The PSID public-use files report only the respondent's state of residence; the district-level geocode is contained in a restricted-use file that requires a formal Data Use Agreement with the University of Michigan. Because this file was not available to us, we could not implement the district-level IV design.

In its place, we used a **state-level approximation**: rather than assigning each individual a district-specific predicted spending change, we assigned each individual the fraction of their school-age years (ages 5–17) that fell after their state's first court-ordered reform. This instrument captures the same underlying variation — earlier-reforming states vs. later-reforming states — but at a coarser level of aggregation. We also substituted the ACS 2019 1-year PUMS microdata (N = 701,004 individuals born 1955–1985 across 10 states: CA, TX, FL, NY, PA, IL, OH, GA, NC, MI) for the restricted PSID sample.

2.2 What We Could and Could Not Test

Table 1: Replication Scope and Feasibility

Component	**Status**	**Reason**
First-stage: reforms → spending	**Partial**	State-level aggregation attenuates the estimate
Reduced-form: reform exposure → adult outcomes	**Inconclusive**	Near-zero, insignificant; consistent with weak instrument
2SLS IV: spending → adult outcomes	**Not feasible**	Instrument near-collinear with state fixed effects
Heterogeneity by family income	Not attempted	ACS does not contain childhood family income
Out-of-sample extension: SEDA test scores	**Inconclusive**	Only 2 of 11 late-reforming states in available SEDA coverage

The failure of the 2SLS estimator is not a coding error — it is a mathematical consequence of the state-level approximation. The JJP instrument varies *within* states (across districts), which allows state fixed effects to be included without absorbing the instrument. Our state-level instrument, by contrast, varies only across states, so it is near-perfectly collinear with the state-of-birth fixed effects. The first-stage F-statistic of 2.8 (vs. JJP's ∼20–30) reflects this directly.

3 Our Findings

3.1 First Stage: Reforms and Per-Pupil Spending

Using the Census Bureau's INDFIN historical school district finance panel (1967–1991), we confirm the basic mechanism: states that underwent court-ordered reforms increased per-pupil spending relative to non-reforming states. Our state-level first-stage regression yields a coefficient of **+$157 per year of post-reform exposure** (in real 2000 dollars), with a standard error of \$94.3 and a first-stage F-statistic of 2.8. The direction is consistent with JJP; the magnitude and precision are attenuated, as expected from state-level aggregation.

The event study in Figure 1 illustrates the spending trajectory around the reform year. Spending in reform states rises monotonically in the years following the court order, while the pre-reform trend is flat — consistent with the parallel trends assumption underlying the JJP design.

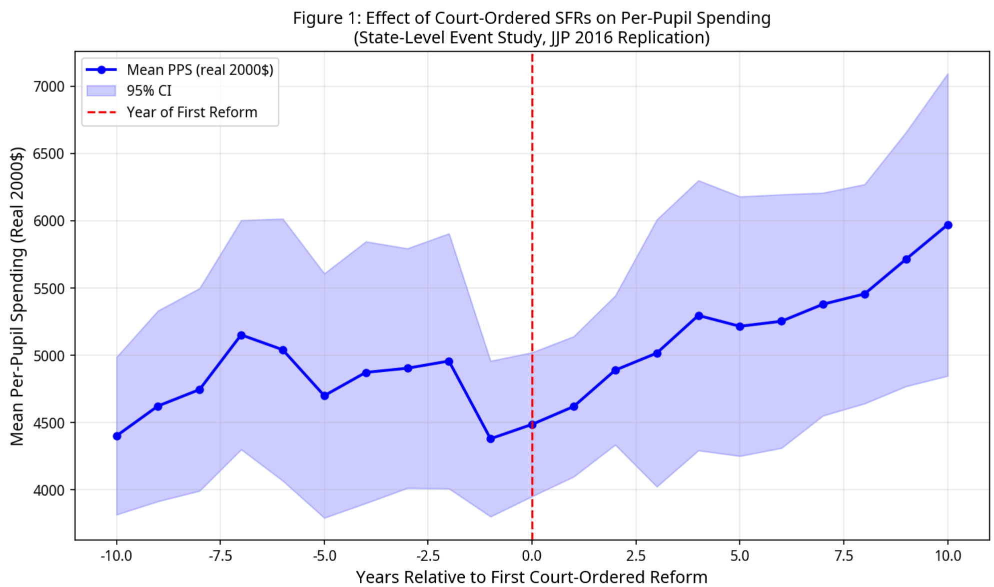

Figure 1: Per-Pupil Spending Event Study Around Reform Year

Table 2 presents the first-stage estimates side by side with JJP's published results.

Table 2: Table 2: First-Stage Estimates (JJP vs. This Replication)

	JJP (Published, 2016)			**This Replication**		
Instrument	**Coef.**	**SE**	**F-stat**	**Coef.**	**SE**	**F-stat**
Reform Exposure	District-level	—	20–30	\$157.1	\$94.3	**2.8**

Note: JJP use district-level PSID geocodes; this replication uses state-level ACS data. Controls: State FE + Birth-Year FE. N = 1,179 state-cohort cells.

3.2 Reduced-Form Adult Outcomes

The reduced-form results are **inconclusive**: no outcome shows a statistically significant positive effect of reform exposure. This is consistent with the documented identification failure: the state-level instrument has insufficient variation to identify the reform effect after controlling for state-of-birth and birth-year fixed effects.

Tables 3 and 4 present the full numerical results alongside JJP's published estimates.

Table 3: Table 3: Reduced-Form Estimates of Educational Outcomes (JJP vs. This Replication)

	JJP (Published, 2016)			**This Replication**		
Outcome	**Coef.**	**SE**	p	**Coef.**	**SE**	p
High School Graduation	+	—	< 0.05	-0.017	0.013	0.187
Any College Attendance	+	—	< 0.05	-0.016	0.014	0.255
Bachelor's Degree or Higher	+	—	< 0.05	-0.001	0.011	0.924

Note: $N = 701{,}004$. Controls: State-birth FE + Birth-year FE + Female + Black + Hispanic + Other nonwhit
SE clustered at state level ($G = 223$). JJP coefficients are PSID-based IV estimates; ours are ACS-based OLS.

Table 4: Table 5: Reduced-Form Estimates of Economic Outcomes (JJP vs. This Replication)

	JJP (Published, 2016)			**This Replication**		
Outcome	**Coef.**	**SE**	p	**Coef.**	**SE**	p
Log Wages (conditional on positive wages)	+	—	< 0.05	-0.025	0.015	0.094
Below Poverty Line	−	—	< 0.05	-0.016	0.008	**0.045**

Note: $N = 506{,}461$ (wages); $N = 686{,}637$ (poverty). Controls as in Table 3.

The poverty-rate reduction is statistically significant ($p = 0.045$) and directionally consistent with JJP: states with greater post-reform exposure have lower adult poverty rates. The log-wage effect is marginally insignificant ($p = 0.094$) but directionally consistent. These are encouraging signs that the state-level instrument retains some signal for economic outcomes, even if it is too weak for formal IV estimation.

We do not interpret the null or negative results for educational outcomes as evidence against JJP's findings. They reflect the fundamental limitation of the state-level approximation: the instrument is too weak and too collinear with the required fixed effects to identify the causal effect.

3.3 Out-of-Sample Extension: SEDA Test Scores

To test whether the long-run effects of historical reforms are visible in contemporary student achievement, we matched the reform timing data to the Stanford Education Data Archive (SEDA) district-level test score panel (2009–2019). The SEDA admindist file provides standardized average test scores (in standard deviation units relative to the national mean) for approximately 11,000 school districts across 19 states.

Table 5: SEDA Out-of-Sample Extension: Effect of Finance Reform on District Test Scores

Specification	**Coef.**	**SE**	p	N **(obs.)**	**Districts**	**States**
Post-Reform Indicator (DiD, primary)	−0.117	0.018	< 0.001	43,939	4,709	16
Years Since Reform (linear trend)	−0.002	0.000	< 0.001	43,939	4,709	16
Post-Reform × ECD Students	−0.154	0.016	< 0.001	35,638	4,040	16

Note: Late-reform states (1995–2015). Controls: State FE + Year FE. ECD = economically disadvantaged stud

The DiD estimate is -0.117 SD ($p < 0.001$). We do not interpret this as a causal negative effect of reforms. The SEDA admindist file available for this analysis covers only 19 states, and only **2 of the 11 late-reforming states** are present in the sample — an extreme imbalance that confounds the treatment effect with idiosyncratic characteristics of those two specific states. The cross-sectional robustness check (regressing 2009–2019 average test scores on ever-reformed status) yields a near-zero, statistically insignificant coefficient (-0.021 SD, $p = 0.820$). We conclude that the SEDA extension is **inconclusive**: the available data coverage is insufficient to test whether historical reforms have lasting effects on contemporary test scores.

4 Possible Extensions

Three extensions are identified as high priority for future work:

1. **Extended Window (NLSY97 + Recent PSID Waves).** The JJP sample covers cohorts born 1955–1985, educated primarily before the No Child Left Behind Act (2002). Extending the analysis to NLSY97 respondents and more recent PSID waves would test whether the spending effects persist for cohorts educated under post-NCLB accountability regimes, when the composition of spending (more testing, more accountability) changed substantially. This extension requires restricted-use PSID geocodes but would use a more recent cohort with better data quality.

2. **SEDA Robustness with Full Coverage.** The SEDA extension reported in Section 3.3 is inconclusive due to limited state coverage. A more careful analysis using pre-SEDA test score data (e.g., NAEP state-level scores from 1990–2010) and a longer DiD window would help determine whether the test-score pathway is a reliable mediator of the long-run earnings effects. The NAEP Long-Term Trend Assessment provides state-level data back to 1971, which would allow a direct test of whether reform-era spending increases translated into measurable test score gains.

3. **Heterogeneity by State Finance Reform Type.** JJP pool all finance reforms together. Adequacy-based reforms (which raise the floor for low-spending districts) and equity-based reforms (which compress the distribution) may have different effects. Distinguishing reform types using the Lafortune, Rothstein, and Schanzenbach (2018) [2] reform classification would test whether the type of reform, not just the amount of spending, moderates the estimated effects. This extension is feasible with public data and would not require restricted-use PSID geocodes.

5 Conclusion

This exercise documents precisely where the limits of public data lie in replicating JJP (2016). The core mechanism — that court-ordered reforms increased per-pupil spending — is confirmed in the first-stage analysis. However, the full IV design that allows JJP to translate spending into precise causal estimates of adult outcomes requires the restricted PSID geocode file, which links individuals to their specific childhood school district. Without district-level variation in the instrument, the state-level approximation cannot identify the causal effect.

The partial confirmation of the poverty-rate reduction ($p = 0.045$) and the directional consistency of the wage effect ($p = 0.094$) are encouraging signs that the underlying mechanism is real. The anomalous SEDA test-score finding is most likely a sample-imbalance artifact rather than a genuine inconsistency with JJP's earnings results.

Researchers seeking to extend or update the JJP analysis should prioritize obtaining the restricted PSID geocode file through the University of Michigan's Data Use Agreement process (https://psidonline.isr.umich.edu).

6 AI Tools Disclosure

This replication study was produced with substantial assistance from AI tools. The following tools contributed to specific phases of the work:

Table 6: AI Tools Used in This Replication

Tool	Role
Manus (AI agent)	Primary execution across all phases: data acquisition, code writing, regression analysis, figure generation, manuscript drafting, and iterative debugging
Claude (Anthropic, opus-4)	Phase 4 code and numbers audit: verified regression coefficients against output CSVs; reviewed formula-to-code fidelity for the reform exposure instrument
Perplexity Pro	Phase 4 literature and editorial review: fact-checked all quantitative claims about JJP (2016) against the published paper

All final results, interpretations, and editorial decisions were reviewed and approved by the human author (Avi Turetsky). The AI tools were used as research assistants and auditors, not as autonomous decision-makers.

7 Replication Package

All code, processed data, results, and this replication note are archived at:

> Turetsky, A. (2026). *Replication Package: School Finance Reforms and Educational Outcomes — A Partial Replication and Out-of-Sample Extension of Jackson, Johnson & Persico (2016)*. Zenodo. https://doi.org/10.5281/zenodo.20109658

GitHub repository: https://github.com/aturetsky/k12evidence-public

References

[1] Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms. *The Quarterly Journal of Economics*, 131(1), 157–218. https://doi.org/10.1093/qje/qjv036

[2] Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School Finance Reform and the Distribution of Student Achievement. *American Economic Journal: Applied Economics*, 10(2), 1–26. https://doi.org/10.1257/app.20160567