Schooling and Labor Market Consequences of School Construction in Indonesia

A Reply to Roodman (*Journal of Comments and Replications in Economics, 2026)*

Esther Duflo[*], Allan Hsiao[†], Mikey Jarrell[‡], and Nathan Lazarus[§]

Journal of Comments and Replications in Economics, Volume 5, 2026-4, DOI: *10.18718/81781.57*

Please Cite As: Duflo, E., Hsiao, A., Jarrell, M., and Lazarus, N. (2026). Schooling and Labor Market Consequences of School Construction in Indonesia: A Reply to Roodman (2026). *Journal of Comments and Replications in Economics*, Vol.5 (2026-6). DOI: 10.18718/81781.57

[*]Department of Economics, Massachusetts Institute of Technology. eduflo@mit.edu
[†]Department of Economics, Stanford University. ajhsiao@stanford.edu
[‡]Department of Economics, UC San Diego. mjarrell@ucsd.edu
[§]Department of Economics, Massachusetts Institute of Technology. nlazarus@mit.edu

Received March 2, 2026; Published April 07, 2026.

Duflo (2001) studies the effects of INPRES, a massive primary school construction project undertaken by the government of Indonesia in the 1970s and 1980s. Using a difference-in-differences (DID) instrumental variables (IV) specification, she finds that school construction boosted educational attainment and wages, implying a return to a year of schooling of around 7%—a number that is consistent with both the contemporaneous (IV) evidence from the United States and also more recent estimates from across the developing and developed world (Card 2001; Patrinos and Psacharopoulos 2020; Gunderson and Oreopolous 2020).

Roodman (2026) reexamines those findings. The contributions include adding more years of data, tracking down the original school construction data, adding corrections to standard errors that have become standard in the literature, and offering new specifications. He finds that the original results largely hold, with less precision than previously thought—with one important exception, which we discuss below.

Before discussing Roodman's findings in detail, some context may be helpful. Roodman began this replication exercise in 2022, and was kind enough to alert Duflo of his progress. Jarrell and Lazarus (2023) replicated and made detailed comments on an earlier Roodman draft (public). Parallel to Roodman, Hsiao (2024) conducted his own replication; like Roodman, Hsiao dipped back into the original data and also extended the sample using 2011–2014 survey waves. This note takes stock of where we have landed, although we would like to note that there are limits to how much we can learn from this exercise. To his credit, Roodman incorporated the Jarrell and Lazarus feedback in subsequent drafts: correcting errors, dropping misguided specifications, etc. The result is that the current draft of Roodman (2026) differs dramatically from its predecessors.

But it is important that replication exercises not follow the stopping rule of "stop when the original result has been debunked," which would create a similar bias in replications to the "file drawer" problem in empirical research. The largest critique of Roodman (2026)—that the results of Duflo (2001) hinge on the inclusion of one particular control—was scarcely mentioned in his first six drafts, with merely an aside calling the controls unmotivated. Instead, other critiques were highlighted, which have since been largely abandoned. Even if this surviving critique were compelling—with which, as we argue below, we disagree—the reader would be justified in being wary of, as Roodman put it in an April 2025 draft, this particular form of "reverse p-hacking,"[1] which may undermine his quest to "reduce researcher degrees of freedom."

The remainder of this note presents replication results in the original dataset (Section 1) and in extended datasets (Section 2), highlighting where we disagree with Roodman's recommendations.

1 *Replicating the original results*

Roodman (2026) corrects data entry errors, clusters standard errors by district (as already suggested by others in the literature), and attempts to apply weak instrument–robust inference tools (Anderson and Rubin 1949). These have minor effects on the results: the point estimates barely move, although some specifications lose significance. We endorse these updates, especially the corrections to the school construction data. By reconstructing this dataset from scratch, Roodman (along with Hsiao (2024), who followed Roodman in doing the same thing) makes a significant contribution considering the amount of follow-on work by economists studying the INPRES shock.[2]

We replicate Roodman's updates using the Duflo (2001) specifications and find nothing to argue about. The wide Anderson–Rubin confidence intervals point to a weak first stage, particularly in the specifications in which many instruments are used.[3] This should come as no great surprise, since the IV literature has made great progress since Duflo (2001) was published.

Roodman then proposes three new specifications. (1) He suggests that survey weights, which Duflo (2001) omits from her regression specifications, should be considered for inclusion even though doing so reduces precision. (2) He uses a two-piece linear spline model to estimate a change in slope rather than a change in levels; doing so controls for pre-trends but also appears to come at a

1. He wrote, "I dropped most discussion of no-controls results in early drafts because it did not seem relevant for kink-based inference. That proved wrong in practice. And on reflection, it seemed appropriate to highlight the influence of a specification choice not motivated in text. That explanation notwithstanding, readers can rationally put some weight on the theory that by expanding this discussion in the course of debate, I have reverse p-hacked."

2. The list of papers that rely on INPRES as a shock to education is long and growing (Duflo 2004b; Breierova and Duflo 2004; Martinez-Bravo 2017; De Chaisemartin and D'Haultfœuille 2018; Porzio and Santangelo 2019; Mazumder, Rosales-Rueda, and Triyana 2019; Ashraf et al. 2020; Bazzi, Hilmy, and Marx 2020; Akresh, Halim, and Kleemans 2021; Hsiao 2024; Gethin 2025).

3. Although Anderson–Rubin confidence intervals tend to be conservative, particularly with multiple instruments (see, e.g. Mikusheva and Sun (2024)), this fact is also visible in the effective F statistics.

cost to power. (3) He suggests that the number of children in a district, which Duflo (2001) includes as a control, should be omitted in at least some specifications, since it is an important source of variation in the number of schools per capita. We address these one by one.

1.1 Weights

Inclusion of survey weights is debatable. In the context of an IV strategy, it is not necessary to weight the regression by survey weights to recover a local average treatment effect. We should use the variation we do have optimally, which is what OLS and 2SLS are meant to do.[4] Weighting is appropriate for summary statistics—which is why Duflo (2001) does so in the descriptive tables (e.g., her Tables 1 and 3).

In IV regressions, we should expect including survey weights to decrease power without changing the point estimates, which is indeed what we see in the results that Roodman (2026) reports. We appreciate his inclusion of the tests for whether the less-powered weighted estimates rule out the unweighted ones. Roodman's example of where the exclusion of weights could introduce bias due to migration across provinces affecting the composition of survey respondents is plausible, but we think this bias will be quantitatively small (consistent with Roodman's results in previous drafts from "weight association tests").

1.2 Spline

The linear spline model is an interesting suggestion, but may come at the cost of some power. Under the null hypothesis, the pre-trend is zero, and we lose a lot of power trying to estimate a nonzero trend. This is similar to the approach of Rambachan and Roth (2023). An alternative approach to this challenge would have been to apply the corrections for pre-trends suggested by Freyaldenhoven, Hansen, and Shapiro (2019). In either case, adding the spline term should decrease power without changing the point estimates, which is indeed what we are seeing in the regressions reported by Roodman. Given the lack of an observable pre-trend in the Duflo (2001) specifications, we don't

4. This was in fact the subject of a favorite exam question of Jerry Hausman for graduate students at MIT.

think it is important for the point estimates.

1.3 Controls

We strongly disagree with Roodman's suggestion that the correct specification ought to exclude a control for the number of children in the district. To see why, it's helpful to view Figure (1), which plots the number of INPRES schools received by districts per child as a function of total child population.

The downward trend is striking. More populous districts received fewer schools per capita on average—districts with less than 50,000 children received nearly twice as many schools per capita as those with 200,000 or more—even though the assignment rule was, on paper, population-agnostic. Table 2 in Duflo (2001) shows this: the coefficient in a log-log specification of schools built on child population is 0.78. As discussed in Duflo (2001), schools were supposed to be allocated according to the fraction of children *not enrolled* (see Duflo (2001) Table 2), but there was substantial residual variation after controlling for enrollment rates. This bias towards school construction in less populated areas resulted in proportionally more schools being assigned to smaller (more rural) districts and fewer to large (urban, often Javanese) districts.

This variation in school construction (and later, potentially, in education and wages) due to the direct effects of regency population is simply not variation we are interested in using. Therefore, given the option to do so, it should be controlled away. Nevertheless, Roodman argues against this control.

> "However, it is not obvious that controlling away a known correlate of treatment improves estimation, since the treatment variation that remains is not necessarily more exogenous than that expunged. One can construct stories of confounding for any nonrandom component of treatment, such as the time-varying effects of the baseline primary school attendance rate, which are also controlled for in the intermediate and full control sets. Duflo (2004a) argues that in Duflo (2001), 'identification is made possible because the

> allocation rule for schools is known.' That point loses some force when one controls away the large, known components of allocation. For the reasons above, it is informative and minimally arbitrary to check the effect of dropping the controls based on population."

This is a misunderstanding of the empirical strategy. Not controlling for a variable that is correlated with both the treatment and the outcome of interest is a misspecification, resulting in omitted variable bias. If, for example, wages were growing especially quickly in Java over this time, then the estimated treatment effect would be downwardly biased.

Our Figure (2) shows that pre-trends differ by the size of the population of children in a district: for each year-of-birth cohort, relative to the 1950 cohort (which serves as the omitted group), it plots the trend in coefficients from a regression of years of schooling on the number of children in a district. The upward trend in education among wage earners indicates that in the treated cohorts (born around 1970), individuals in more populated areas had relatively greater educational attainment, while this difference was more muted in the control cohorts (born around 1960). Figure (2) also plots the same trend for wages, which reveals increasing coefficients, implying that higher population is more strongly correlated with higher wages in the younger cohorts. A similar pattern is seen in the later SUSENAS data.

These differential trends associated with child population appear to persist throughout the pre-periods and into the treatment periods (without a trend break), making it clear that child population is likely to be a confounding variable and, therefore, should be controlled for. The opposite extreme, if we believed the variation in school building from population were worth using, would be to run an instrumental variables regression of wages on schooling and instrument for schooling with population: the exclusion restriction—that population can only affect wage growth via its effect on school construction—would clearly be violated.

It should therefore come as no surprise that omitting this control makes a marked difference. Our Table (1) illustrates this. In Column (1) of Panel E, we replicate the first-stage regression from the single-instrument specification in Duflo (2001) (see her Equation (1))—i.e., the effect of INPRES on schooling among the population of males who report earning wages in the 1995 SUPAS—while incorporating the Roodman (2026) corrections that we endorse. Omitting the child population control wipes out the first stage completely. But in Column (2), when the control is included, a useful first stage appears.

This explains the preferred specification of Duflo (2001), which controls for total number of children and for elementary school enrollment rates, precisely because they are known components of the allocation rule. The resulting identification assumption is that, absent INPRES, two places that have the same child population size and baseline enrollment rate but received a different number of INPRES schools per child would have followed parallel trends in terms of schooling and wages.

Roodman argues that there may be other omitted variables that are correlated with both the number of INPRES schools received per child conditional on child population and enrollment rate and correlate with growth in schooling and wages. There may be such omitted variables; one of them is other programs following the same rules, which is why Duflo (2001) controls for the water and sanitation program, or opposite rules (for example, we see that the increase in secondary education follows the opposite pattern, perhaps because they did build more of those schools in regions with high enrollment rates). This is ultimately the identifying assumption—we can never rule out that variation from a "natural" experiment in fact suffers from unobserved confounds like these, and the trend break and the various specification checks are meant to provide some amount of confidence.

In retrospect, Duflo (2001) could in fact have controlled for child population more flexibly to exploit only the variation conditional on the number of children. We show how this influences results in Panel A of Tables (4) and (5), where we replicate the IV results of Duflo (2001) while adding an additional control: dummy variables corresponding to each decile of the distribution of number of children in a district, which we interact with year of birth dummies. This produces results that are more precise and more stable across all the other controls.

2 *Extending the datasets*

Many years have passed since the original paper was published in 2001. Roodman (2026) extends the analysis to later surveys of the same cohorts, adding the 2005 SUPAS and the 2011–14 and

2017–19 rounds of the SUSENAS to the 1995 SUPAS used by Duflo (2001). Hsiao (2024) does the same as a preliminary exercise as part of his study of migration as a mechanism to explain the returns to education. Jarrell and Lazarus (2023) independently replicate their results. Table (2) compares the four sets of results for the IV estimates of the returns to education: a replication of Duflo's 2001 results using the 1995 SUPAS (Panel A), and three sets of results using the 2011–14 waves of the SUSENAS. The SUSENAS results differ in how this new data is cleaned: Panel B uses our preferred cleaning approach (which we refer to as "Jarrell–Lazarus"), Panel C uses the approach employed in Hsiao (2024), and Panel D uses the approach employed in Roodman (2026).[5]

The new data decisively replicates the Duflo (2001) results, which seem to be fairly robust across different cleaning conventions. Roodman, like Hsiao, finds that the impacts of INPRES persisted even as these cohorts aged through the decades. Deploying the main specifications of Duflo (2001) to this new data, Roodman finds positive and significant effects of INPRES on schooling and wages. Even after applying Roodman's corrections to this follow-up data—clustering, correcting data errors, and calculating Anderson–Rubin weak instrument–robust confidence intervals—the estimated returns to schooling are large and significant, as are the first stage and the reduced form results.

Not surprisingly (and for the same reason as before), omitting the number of children control also erases any trace of statistical significance from the IV results in this new data, as can be seen in Column (1) across Tables (2) and (3). As in the SUPAS data, these IV results are not identified because there is no first stage in the wage earner sample without the control. Again, there is confounding due to differential trends in populous and less populous regencies.

5. The differences between the Jarrell–Lazarus cleaning and Hsiao's cleaning stem almost entirely from two choices: Hsiao truncates years of education at 12 (to study the effects of primary and secondary schooling), and Hsiao reconstructed the enrollment control from an alternative source (the 1971 census as opposed to the Ministry of Education and Culture). We show both because using the non-truncated education and the Ministry of Education and Culture enrollment data hew most closely to the specifications in the original paper, but the alternative education coding and enrollment data are reasonable specification choices made by Hsiao in his replication and extension. The divergences between the Jarrell–Lazarus cleaning and Roodman's are documented in more detail at this link. They grew much more similar after our prior back-and-forth, and we both benefited from each other's comments on data construction details. Those that remain are largely attributable to Roodman's choice to award years of schooling to students who attempted but did not complete the year; Roodman counting undergraduate degrees as five years of schooling while we count them as four; and a few errors we believe we've identified in the crosswalk Roodman uses to handle changing regency definitions. We were not able to perfectly replicate the numbers from Roodman (2026) because his public code did not show his full process of matching regencies of birth among respondents to the 2011–2014 SUSENAS to 1975 regency definitions, but we believe we did this in a very similar way, using crosswalks from prior versions of Roodman's code, and the differences between our replication of his values and the values in his paper are quite small.

Also as in the SUPAS data, when we add the flexible population controls to allow for different trends across deciles in Panels B–D of Tables (4) and (5), we find that this extra control helps with precision and stability across specifications, and does not otherwise significantly alter the results—as was the case for the SUPAS results in Panel (A) of the same tables.

3 Conclusion

Science keeps progressing. Relative to 2001, we now have more data on children who benefited from the INPRES program, more econometrics knowledge, and new techniques for applying that knowledge. It is exciting to see that the Indonesia experiment continues to be interesting to scholars. All in all, we believe that Roodman's paper (except for the one misspecification) reinforces the original conclusions that the program increased schooling and earnings. The robustness of the finding to the inclusion of later survey years is particularly interesting.

Roodman does present us with the opportunity to highlight some important limitations of Duflo (2001). First, none of the standard errors were clustered (not even at the level of year of birth interacted with district of birth, let alone at the district level). This results in overstated precision. Second, the first stage suffers from a weak instruments problem, especially with multiple instruments. Third, state-of-the-art specifications for difference-in-differences estimates with a continuous treatment were not used (Freyaldenhoven, Hansen, and Shapiro 2019; Callaway, Goodman-Bacon, and Sant'Anna 2024).

Nevertheless, in our view, the substantive conclusion that the INPRES schools led to more schooling and that schooling increases wages was likely accurate, especially since a substantive literature confirms these sorts of results. Despite these limitations, the natural experiment created by INPRES can still be used to look at the impact of education on other important outcomes. In light of the recent econometric developments in the IV literature, we encourage future researchers to seek as many waves of these surveys as possible to increase precision, to focus on more reduced-form estimates, or to make use of structural specifications where the school construction variable itself is used as a source of variation.

It is important to note that we don't see either this paper or the subsequent literature as being anywhere close to the last word on the subject. There is very little experimental work on returns to education—for the good reasons that it takes a long time and requires large samples. And there is very little work of any kind on how to think about general equilibrium effects. A lot remains to be learned, and we hope scholars will continue to bring new data and experiments (natural or not) to the table.

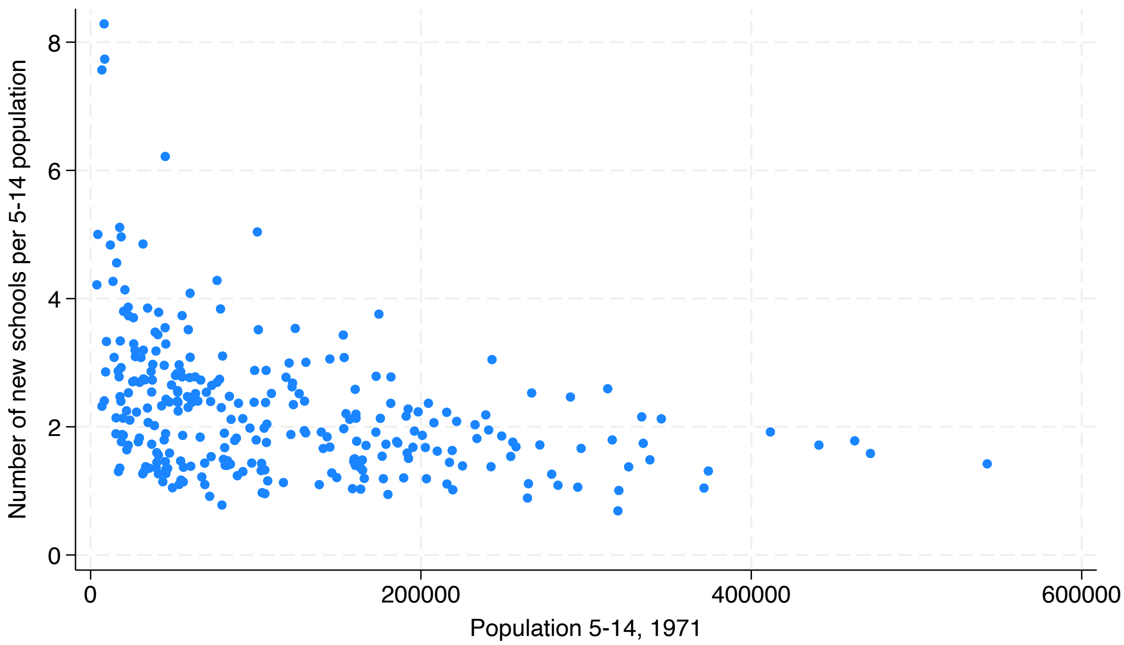

Figure 1: INPRES School Construction Per Capita as a Function of Number of Children in a District

Note: This plots the number of INPRES schools by regency after applying Roodman's corrections to the original data against child population.

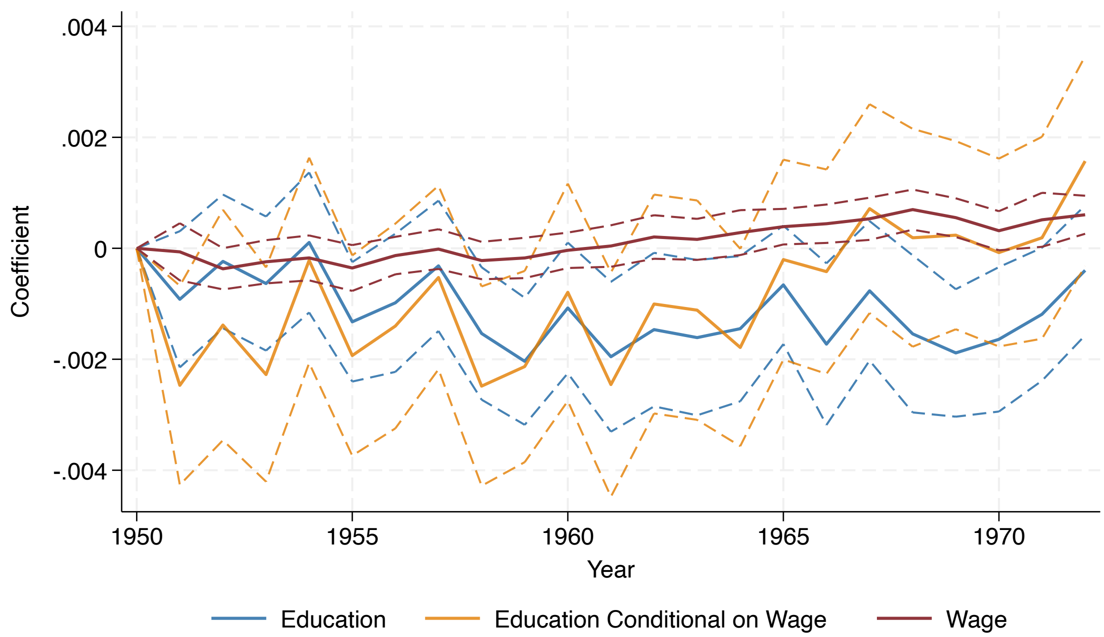

Figure 2: Effect of Number of Children in a District on Schooling and Wages by Birth Cohort (SUPAS 1995 Data)

Note: This plots regressions of outcomes (log hourly wage and years of education) on number of children interacted with birth year. The 1950 interaction is the omitted category, so coefficients are relative to the effect of number of children in that year. All regressions have fixed effects for birth province and birth year. Education conditional on wage is the regression with years of education as an outcome among only wage earners. The units of the education result are years, and the units of wages are log wages; they are not standardized to be comparable.

Table 1: Single-IV First Stage (Years of Education)

	(1)	(2)	(3)	(4)	(5)
Panel A: Supas (1995) (Full Sample)					
	0.100	0.119	0.082	0.094	0.139
	(0.041)	(0.042)	(0.041)	(0.043)	(0.043)
Obs	78,470	78,470	78,470	73,153	78,470
Panel B: SUSENAS (2011–14), Jarrell–Lazarus Cleaning (Full Sample)					
	0.074	0.116	0.026	0.035	0.165
	(0.036)	(0.038)	(0.034)	(0.035)	(0.037)
Obs	301,034	301,034	301,034	295,627	301,034
Panel C: SUSENAS (2011–14), Roodman's Cleaning (Full Sample)					
	0.070	0.101	0.016	0.025	0.143
	(0.033)	(0.035)	(0.031)	(0.032)	(0.035)
Obs	300,147	300,147	300,147	294,565	300,147
Panel D: SUSENAS (2011–14), Hsiao's Cleaning (Full Sample)					
	0.089	0.131	0.072	0.085	0.180
	(0.036)	(0.037)	(0.037)	(0.037)	(0.036)
Obs	301,033	301,033	301,033	301,033	301,033
Panel E: Supas (1995) (Wage Earners)					
	0.029	0.159	0.132	0.163	0.175
	(0.054)	(0.058)	(0.057)	(0.063)	(0.063)
Obs	31,061	31,061	31,061	28,183	31,061
Panel F: SUSENAS (2011–14), Jarrell-Lazarus Cleaning (Wage Earners)					
	–0.016	0.090	0.028	0.057	0.128
	(0.044)	(0.046)	(0.048)	(0.047)	(0.048)
Obs	113,736	113,736	113,736	110,836	113,736
Panel G: SUSENAS (2011–14), Roodman's Cleaning (Wage Earners)					
	–0.022	0.072	0.019	0.046	0.101
	(0.043)	(0.045)	(0.047)	(0.046)	(0.047)
Obs	112,953	112,953	112,953	109,933	112,953
Panel H: SUSENAS (2011–14), Hsiao's Cleaning (Wage Earners)					
	0.024	0.118	0.081	0.106	0.167
	(0.039)	(0.040)	(0.042)	(0.041)	(0.042)
Obs	112,877	112,877	112,877	112,877	112,877
Birthyear × Number of Children in 1971		✓	✓	✓	✓
Birthyear × Decile of Number of Children in 1971					✓
Birthyear × Enrollment Rate in 1971			✓	✓	
Birthyear × Water and Sanitation Program				✓	

Note: Each column reports first stage estimates of the effect of the instrument (young cohort interacted with program intensity) on years of education. Panels A–D use the full sample; Panels E–H restrict to wage earners. In parentheses are standard errors clustered at the region of birth.

Table 2: Returns to Schooling Estimated using Different Datasets (Single Instrument)

	(1)	(2)	(3)	(4)
Panel A: 1995 SUPAS				
	–0.499	0.118	0.090	0.136
	(1.150)	(0.058)	(0.072)	(0.068)
AR CI	$(-\infty, \infty)$	[–0.03, 0.32]	[–0.19, 0.42]	[–0.01, 0.49]
MOP F	0.28	7.38	5.19	6.76
Obs	31,061	31,061	31,061	28,183
Panel B: Jarrell–Lazarus Cleaning of 2011–14 SUSENAS				
	0.284	0.269	0.690	0.402
	(0.616)	(0.107)	(0.973)	(0.255)
AR CI	$(-\infty, \infty)$	[–0.49, 14.43]	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	0.13	3.90	0.34	1.44
Obs	113,736	113,736	113,736	110,836
Panel C: Hsiao Cleaning of 2011–14 SUSENAS				
	–0.038	0.229	0.283	0.241
	(0.335)	(0.076)	(0.132)	(0.093)
AR CI	$(-\infty, \infty)$	[0.09, 0.62]	[–2.09, 55.18]	[0.08, 0.86]
MOP F	0.50	9.40	4.18	7.15
Obs	112,877	112,877	112,877	112,877
Panel D: Roodman Cleaning of 2011–14 SUSENAS				
	0.280	0.307	0.954	0.473
	(0.412)	(0.145)	(2.027)	(0.365)
AR CI	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	0.26	2.60	0.17	0.98
Obs	112,953	112,953	112,953	109,933
Birthyear × Number of Children in 1971		✓	✓	✓
Birthyear × Enrollment Rate in 1971			✓	✓
Birthyear × Water and Sanitation Program				✓

Note: Each column reports an estimate of the returns to schooling using an instrumental variables specification in which years of schooling is instrumented for by the interaction between INPRES school construction intensity and an dummy variable for being born in the "young" (i.e., treated) cohort. The coefficients represent the effect of one year of schooling on log hourly wages. In parentheses are standard errors clustered at the level of the district of birth. AR CI reports 95% Anderson–Rubin confidence intervals from wild bootstrap. MOP F reports the Montiel-Pflueger effective F statistics (Montiel Olea and Pflueger 2013). Like Duflo (2001), Panel A uses the 1995 SUPAS for its outcome data. Panels B–D use the 2011, 2012, 2013, and 2014 waves of the SUSENAS. Panels B, C, and D apply the preferred cleaning approaches of Jarrell and Lazarus (2023), Hsiao (2024), and Roodman (2026). All specifications include year of birth fixed effects and region of birth fixed effects. Column (1) omits all other controls. Columns (2)–(4) add three different district-level control variables, all interacted with year of birth dummies: child population, primary school enrollment rate, and allocation of the INPRES water and sanitation program.

Table 3: Returns to Schooling Estimated using Different Datasets (Many Instruments)

	(1)	(2)	(3)	(4)
Panel A: 1995 SUPAS				
	0.046	0.108	0.078	0.127
	(0.071)	(0.048)	(0.054)	(0.049)
AR CI	$(-\infty, 1771]$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	0.35	1.56	1.10	1.64
Obs	60,663	60,663	60,663	55,341
Panel B: Jarrell–Lazarus Cleaning of 2011–14 SUSENAS				
	0.079	0.121	0.112	0.112
	(0.038)	(0.045)	(0.056)	(0.054)
AR CI	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	1.34	2.15	0.89	1.17
Obs	189,771	189,771	189,771	185,078
Panel C: Hsiao Cleaning of 2011–14 SUSENAS				
	0.053	0.137	0.154	0.148
	(0.056)	(0.045)	(0.069)	(0.059)
AR CI	$(-\infty, \infty)$	[–0.07, 2.33]	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	1.51	4.35	1.97	2.81
Obs	188,139	188,139	188,139	188,139
Panel D: Roodman Cleaning of 2011–14 SUSENAS				
	0.078	0.115	0.105	0.101
	(0.035)	(0.046)	(0.052)	(0.051)
AR CI	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	1.44	1.80	0.93	1.17
Obs	188,420	188,420	188,420	183,557
Birthyear $\times$ Number of Children in 1971		✓	✓	✓
Birthyear $\times$ Enrollment Rate in 1971			✓	✓
Birthyear $\times$ Water and Sanitation Program				✓

Note: Each column reports an estimate of the returns to schooling using an instrumental variables specification in which years of schooling is instrumented for by the interaction between INPRES school construction intensity and dummy variables for year of birth. The coefficients represent the effect of one year of schooling on log hourly wages. In parentheses are standard errors clustered at the level of the district of birth. AR CI reports 95% Anderson–Rubin confidence intervals from wild bootstrap. MOP F reports the Montiel-Pflueger effective F statistics (Montiel Olea and Pflueger 2013). Like Duflo (2001), Panel A uses the 1995 SUPAS for its outcome data. Panels B–D use the 2011, 2012, 2013, and 2014 waves of the SUSENAS. Panels B, C, and D apply the preferred cleaning approaches of Jarrell and Lazarus (2023), Hsiao (2024), and Roodman (2026). All specifications include year of birth fixed effects and region of birth fixed effects. Column (1) omits all other controls. Columns (2)–(4) add three different district-level control variables, all interacted with year of birth dummies: child population, primary school enrollment rate, and allocation of the INPRES water and sanitation program.

Table 4: Returns to Schooling with Flexible Controls (Single Instrument)

	(1)	(2)	(3)	(4)
Panel A: 1995 SUPAS				
	0.144	0.143	0.124	0.149
	(0.055)	(0.056)	(0.078)	(0.077)
AR CI	[0.03, 0.39]	[0.02, 0.40]	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	7.65	7.53	3.44	4.29
Obs	31,061	31,061	31,061	28,183
Panel B: Jarrell–Lazarus Cleaning of 2011–14 SUSENAS				
	0.212	0.213	0.431	0.348
	(0.059)	(0.059)	(0.317)	(0.177)
AR CI	[0.10, 0.50]	[0.10, 0.49]	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	7.90	8.13	1.06	1.97
Obs	113,736	113,736	113,736	110,836
Panel C: Hsiao Cleaning of 2011–14 SUSENAS				
	0.178	0.177	0.192	0.186
	(0.048)	(0.048)	(0.070)	(0.065)
AR CI	[0.08, 0.33]	[0.08, 0.33]	[0.05, 0.53]	[0.06, 0.46]
MOP F	17.37	17.35	8.70	10.21
Obs	112,877	112,877	112,877	112,877
Panel D: Roodman Cleaning of 2011–14 SUSENAS				
	0.238	0.239	0.555	0.394
	(0.076)	(0.076)	(0.578)	(0.235)
AR CI	[0.11, 0.97]	[0.11, 0.91]	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	5.26	5.42	0.56	1.50
Obs	112,953	112,953	112,953	109,933
Birthyear × Decile of Number of Children in 1971	✓	✓	✓	✓
Birthyear × Number of Children in 1971		✓	✓	✓
Birthyear × Enrollment Rate in 1971			✓	✓
Birthyear × Water and Sanitation Program				✓

Note: Each column reports an estimate of the returns to schooling using an instrumental variables specification in which years of schooling is instrumented for by the interaction between INPRES school construction intensity and a dummy variable for being born in the "young" (i.e., treated) cohort. The coefficients represent the effect of one year of schooling on log hourly wages. In parentheses are standard errors clustered at the level of the district of birth. AR CI reports 95% Anderson–Rubin confidence intervals from wild bootstrap. MOP F reports the Montiel-Pflueger effective F statistics (Montiel Olea and Pflueger 2013). Like Duflo (2001), Panel A uses the 1995 SUPAS for its outcome data. Panels B–D use the 2011, 2012, 2013, and 2014 waves of the SUSENAS. Panels B, C, and D apply the preferred cleaning approaches of Jarrell and Lazarus (2023), Hsiao (2024), and Roodman (2026). All specifications include year of birth fixed effects and region of birth fixed effects. Column (1) omits all other controls. Columns (2)–(5) add four different district-level control variables, all interacted with year of birth dummies: child population (flexibly by decile bins, in which 28–29 of the 282 are assigned to each decile), child population (linearly), primary school enrollment rate, and allocation of the INPRES water and sanitation program.

Table 5: Returns to Schooling with Flexible Controls (Many Instruments)

	(1)	(2)	(3)	(4)
Panel A: 1995 SUPAS				
	0.134	0.133	0.099	0.128
	(0.044)	(0.045)	(0.056)	(0.050)
AR CI	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	1.90	1.83	1.38	1.61
Obs	60,663	60,663	60,663	55,341
Panel B: Jarrell–Lazarus Cleaning of 2011–14 SUSENAS				
	0.103	0.104	0.069	0.082
	(0.032)	(0.032)	(0.049)	(0.048)
AR CI	[–0.62, 0.22]	[–0.56, 0.22]	$(-\infty, \infty)$	$(-\infty, \infty)$
MOP F	3.44	3.51	1.24	1.41
Obs	189,771	189,771	189,771	185,078
Panel C: Hsiao Cleaning of 2011–14 SUSENAS				
	0.108	0.107	0.106	0.105
	(0.031)	(0.032)	(0.046)	(0.044)
AR CI	[–0.13, 0.24]	[–0.13, 0.24]	$(-\infty, 23023]$	[–3.46, 0.28]
MOP F	6.78	6.75	3.32	3.65
Obs	188,139	188,139	188,139	188,139
Panel D: Roodman Cleaning of 2011–14 SUSENAS				
	0.093	0.095	0.054	0.069
	(0.034)	(0.034)	(0.047)	(0.048)
AR CI	[–2.01, 0.21]	[–1.80, 0.21]	$(-\infty, \infty)$	$(-\infty, 17095]$
MOP F	2.72	2.76	1.23	1.38
Obs	188,420	188,420	188,420	183,557
Birthyear × Decile of Number of Children in 1971	✓	✓	✓	✓
Birthyear × Number of Children in 1971		✓	✓	✓
Birthyear × Enrollment Rate in 1971			✓	✓
Birthyear × Water and Sanitation Program				✓

Note: Each column reports an estimate of the returns to schooling using an instrumental variables specification in which years of schooling is instrumented for by the interaction between INPRES school construction intensity and dummy variables for year of birth. The coefficients represent the effect of one year of schooling on log hourly wages. In parentheses are standard errors clustered at the level of the district of birth. AR CI reports 95% Anderson–Rubin confidence intervals from wild bootstrap. MOP F reports the Montiel-Pflueger effective F statistics (Montiel Olea and Pflueger 2013). Like Duflo (2001), Panel A uses the 1995 SUPAS for its outcome data. Panels B–D use the 2011, 2012, 2013, and 2014 waves of the SUSENAS. Panels B, C, and D apply the preferred cleaning approaches of Jarrell and Lazarus (2023), Hsiao (2024), and Roodman (2026). All specifications include year of birth fixed effects and region of birth fixed effects. Column (1) omits all other controls. Columns (2)–(5) add four different district-level control variables, all interacted with year of birth dummies: child population (flexibly by decile bins, in which 28–29 of the 282 are assigned to each decile), child population (linearly), primary school enrollment rate, and allocation of the INPRES water and sanitation program.

References

Akresh, Richard, Daniel Halim, and Marieke Kleemans. 2021. "Long-Term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia," Policy Research Working Papers, https://doi.org/10.1596/1813-9450-9559.

Anderson, Theodore W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *The Annals of Mathematical Statistics* 20 (1): 46–63. https://doi.org/10.1214/aoms/1177730090.

Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena. 2020. "Bride Price and Female Education." *Journal of Political Economy* 128 (2): 591–641. https://doi.org/10.1086/704572.

Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx. 2020. "Religion, Education, and Development," https://doi.org/10.3386/w27073.

Breierova, Lucia, and Esther Duflo. 2004. *The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?* w10513. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w10513.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2024. "Event Studies with a Continuous Treatment." *AEA Papers and Proceedings* 114:601–605. https://doi.org/10.1257/pandp.20241047.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5): 1127–1160. https://doi.org/10.1111/1468-0262.00237.

De Chaisemartin, Clément, and Xavier D'Haultfœuille. 2018. "Fuzzy Differences-in-Differences." *The Review of Economic Studies* 85 (2): 999–1028. https://doi.org/10.1093/restud/rdx049.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91 (4): 795–813. https://doi.org/10.1257/aer.91.4.795.

Duflo, Esther. 2004a. "Scaling up and Evaluation." In *Annual World Bank Conference on Development Economics: Accelerating Development,* edited by Francois Bourguignon, and Boris Pleskovic. The World Bank. ISBN: 978-0-8213-5800-9, https://doi.org/10.1596/0-8213-5800-6.

Duflo, Esther. 2004b. "The medium run effects of educational expansion: evidence from a large school construction program in Indonesia." *Journal of Development Economics* 74 (1): 163–197. https://doi.org/10.1016/j.jdeveco.2003.12.008.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro. 2019. "Pre-Event Trends in the Panel Event-Study Design." *American Economic Review* 109 (9): 3307–3338. https://doi.org/10.1257/aer.20180609.

Gethin, Amory. 2025. "Distributional Growth Accounting: Education and the Reduction of Global Poverty, 1980–2019." *The Quarterly Journal of Economics* 140 (4): 2571–2618. https://doi.org/10.1093/qje/qjaf033.

Gunderson, Morley, and Philip Oreopolous. 2020. "Returns to education in developed countries." In *The Economics of Education,* 39–51. Elsevier. ISBN: 978-0-12-815391-8, https://doi.org/10.1016/B978-0-12-815391-8.00003-3.

Hsiao, Allan. 2024. "Educational Investment in Spatial Equilibrium: Evidence from Indonesia," https://allanhsiao.com/files/Hsiao_schools.pdf.

Jarrell, Mikey, and Nathan Lazarus. 2023. "Schooling and Labor Market Consequences of School Construction in Indonesia: Reply to Roodman (2023)." *Working Paper,* https://github.com/NathanLazarus/ReplyToRoodman/blob/main/DJL_Reply_2023.pdf.

Martinez-Bravo, Monica. 2017. "The Local Political Economy Effects of School Construction in Indonesia." *American Economic Journal: Applied Economics* 9 (2): 256–289. https://doi.org/10.1257/app.20150447.

Mazumder, Bhashkar, Maria Rosales-Rueda, and Margaret Triyana. 2019. "Intergenerational Human Capital Spillovers: Indonesia's School Construction and Its Effects on the Next Generation." *AEA Papers and Proceedings* 109:243–249. https://doi.org/10.1257/pandp.20191059.

Mikusheva, Anna, and Liyang Sun. 2024. "Weak identification with many instruments." *The Econometrics Journal* 27 (2): C1–C28. https://doi.org/10.1093/ectj/utae007.

Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3): 358–369. https://doi.org/10.1080/00401706.2013.806694.

Patrinos, Harry Anthony, and George Psacharopoulos. 2020. “Returns to education in developing countries.” In *The Economics of Education,* 53–64. Elsevier. ISBN: 978-0-12-815391-8, https://doi.org/10.1016/B978-0-12-815391-8.00004-5.

Porzio, T., and G. Santangelo. 2019. “Does Schooling Cause Structural Transformation?” Number: 1925, *Cambridge Working Papers in Economics,* https://ideas.repec.org//p/cam/camdae/1925.html.

Rambachan, Ashesh, and Jonathan Roth. 2023. “A More Credible Approach to Parallel Trends.” *Review of Economic Studies* 90 (5): 2555–2591. https://doi.org/10.1093/restud/rdad018.

Roodman, David. 2026. “Schooling and Labor Market Consequences of School Construction in Indonesia: A Replication Study of Duflo (American Economic Review, 2001).” *Journal of Comments and Replications in Economics* 5 (5). https://doi.org/10.18718/81781.56.