Public Primary School Expansion, Gender-based Crowding Out, and Intergenerational Educational Mobility

Ahsan, Md Nazmul and Shilpi, Forhad and Emran, Shahe

Saint Louis University, World Bank, IPD, Columbia

4 March 2023

PUBLIC PRIMARY SCHOOL EXPANSION, GENDER-BASED CROWDING OUT, AND INTERGENERATIONAL EDUCATIONAL MOBILITY

Md. Nazmul Ahsan, Saint Louis University[1]

M. Shahe Emran, IPD at Columbia University

Forhad Shilpi, DECRG, World Bank

First Draft: October 15, 2021
This Version: March 4, 2023

ABSTRACT

From 1965 to 1985, the number of schools doubled in developing countries, but little is known about their impacts on intergenerational educational mobility. We study the effects of 61,000 public primary schools constructed in the 1970s in Indonesia on intergenerational educational mobility, using full-count census data and a credible identification strategy. The educational mobility curve is concave in most of the cases, and school expansion reduced the degree of concavity. Evidence from a DiD design on primary completion suggests contrasting effects across the distribution: relative mobility improved irrespective of gender in the uneducated households, but it worsened in the highly educated households. For completed years of schooling, there are striking gender differences, with strong effects on sons, but no significant effects on girls. This surprising finding reflects an unintended bottleneck at the secondary schooling level which created a fierce competition among the Inpres primary graduates. The girls suffered an 8.5 percentage points decline in the probability of completing senior secondary schooling, while the boys reaped a 7.7 percentage points gain. The gender-based crowding out occurred across the board, suggesting mechanisms unrelated to family background such as low labor market returns for girls and gender norms in a patrilineal society. Available evidence on returns to education of girls rejects a labor market based explanation. We test and find evidence consistent with gender norms as a mechanism by exploiting data from the "Matrilineal island" West Sumatra. In West Sumatra, girls are not crowded out at the secondary level, instead the boys face significant crowding out.

Key Words: Public Schools, Intergenerational Mobility, Education, Heterogeneous Impacts, Most Disadvantaged Children, Social Norms, Indonesia, West Sumatra, Patrilineal, Matrilineal, Unintended Bottleneck, Gender-based Crowding Out, Full Count Census Data

JEL Codes: I24, J62, J16, O20

---

[1]Emails for correspondence: nazmul.ahsan@slu.edu (Md Nazmul Ahsan), shahe.emran.econ@gmail.com (M. Shahe Emran), fshilpi@worldbank.org (Forhad Shilpi). We are grateful to Margaret Triyana, Jan Stuhler, Daniel Suryadarma, and Hanchen Jiang for extensive comments on an earlier version. We would also like to thank Jere Behrman, Marieke Kleemans, Tatjana Kleinberg, Arya Gaduh, Pablo Mitnik, Cynthia Bansak, M. Najeeb Shafiq, and participants at the GLO Global Conference 2022 for discussions and/or comments on an earlier draft. We are grateful to Esther Duflo for generously sharing data on school construction in Indonesia, and Sailesh Tiwari, Imam Setiawan, and Rinku Murgai for help with the census data. *This paper supersedes and supplants an earlier substantially different version titled "Unintended bottleneck and essential nonlinearity: Understanding the effects of public primary school expansion on intergenerational educational mobility".* Standard disclaimers apply.

(1) Introduction

Governments in many developing countries implemented a vast expansion of public schools, especially at the primary level, in the last 60 years. From 1965 to 1985, the number of schools doubled and the number of teachers tripled in developing countries, creating 185 million new school places in two decades (Lockheed et al. (1991)). Primary school infrastructure expansion remains a policy priority in many countries, especially in Africa and South Asia, to provide better access to millions of primary school aged children who remain out of school.[2] There is a large and growing literature analyzing the effects of school expansion on a variety of educational and labor market outcomes, but there is little evidence on its impacts on intergenerational educational mobility. Given an increasing emphasis on equality of opportunity as a salient policy objective (as opposed to equality of outcomes), understanding how such dramatic school expansion affected intergenerational educational mobility in developing countries can be valuable for both the policymakers and donors.[3]

It is often argued that a better access to public schools helps reduce inequality by improving educational mobility of the children from disadvantaged socioeconomic background. The Coleman Report (Coleman et al. (1966)), commissioned by the Unites States Congress, suggested that the expansion of schooling access would reduce educational inequality as children from disadvantaged families gain higher schooling while children from advantaged families hit the ceiling. However, it is also widely recognized that the incidence of public policies such as school construction may not be distributionally progressive or neutral (Becker (1981), World Bank (2006)). The children from advantaged economic background may benefit disproportionately more from a new school because their parents can invest in complementary inputs such as books and private tutors.[4]

[2]According to a recent global estimate, 67 million primary school aged children are out of school, and about 70 percent of them are in Sub-Saharan Africa and South Asia. For more details, see UIS (2022)).

[3]World Bank portfolio of active education projects in 95 developing countries was US $23.61 billion on June 30, 2022 (World Bank Education Fact Sheet, September 14, 2022). 24 percent of the total financing is allocated to primary schools.

[4]There is a substantial literature in sociology on the consequences of *education expansion* where the focus is on the relation between average level of education and educational inequality in a country. There is substantial evidence that the relation between education expansion and educational inequality is inverted-U. It is important to appreciate the difference between our analysis of school expansion which is a supply side intervention, and education expansion which is an equilibrium outcome determined by both school supply and schooling demand.

We provide evidence on the effects of a dramatic expansion of public primary schools in Indonesia in the 1970s on the inheritance of educational inequality across generations.[5] The Sekolah Dasar (SD) Inpres program under the second five year plan constructed more than 61,000 new primary schools and doubled the number of primary schools in five years.[6] Our analysis focuses on two issues: (1) heterogeneity in the effects across the distribution with special reference to the most disadvantaged children (parents with no schooling), and (ii) gender differences in the incidence of the effects. Becker et al. (2015) develop a model where the intergenerational educational mobility equation can be concave or convex implying heterogeneous relative mobility at different levels of a father's education.[7] A concave mobility curve implies that the children of uneducated parents face the lowest relative mobility and raises the possibility of a low education trap. The expansion of primary schools can affect the shape of nonlinearity, implying substantially different effects on relative mobility of the children from different family background.[8]

The effects of school expansion may differ across gender because of social norms such as son preference in a patrilineal society, and differences in the costs of accessing a distant school (World Bank (2018), Tilak (1993), Scott (1985)). Distance hinders the girls more as parents are unwilling to send them to far away schools because of gender norms and safety concerns.[9] If distance to school is a binding constraint on girls' schooling, then we would expect a bigger impact of the Inpres schools on educational opportunities of daughters. When primary school expansion is not backed by similar expansions at the secondary and higher levels, this may create an unintended bottleneck as a large number of primary graduates compete for limited slots at the secondary level. In the face of such a bottleneck, relatively weaker subgroups in a society (e.g., ethnic minority, immigrants) may be crowded out. In a patrilineal society, public

[5]Indonesian government invested US $500 million in school construction in 1973 (Duflo (2001)).

[6]The Inpres school construction has been the focus of a series of interesting studies, see, for example, Pitt et al. (1993), Duflo (2001), Ashraf et al. (2020), Mazumder et al. (2019), Bazzi et al. (2020), Bau et al. (2020).

[7]The concavity arises naturally from diminishing returns to financial investments, and convexity may arise from a variety of sources generating complementarity such as role model and peer effects, and more efficient educational investments by the educated parents.

[8]Taking stock of the economic literature on intergenerational mobility, Cholli and Durlauf (2022) suggest that the "next generation" studies on intergenerational mobility need to go beyond the workhorse linear model and explore the implications of the nonlinearity implied by theory.

[9]Even when the school in the neighboring village is within commuting distance, fear of harassment on the way to school can be an important hurdle for girls.

primary school expansion may result in crowding out of girls at the post primary levels.

For our empirical analysis, we use the full count census data from census 2000 (BPS, Government of Indonesia) and follow closely the difference-in-difference (DiD) strategy developed by Duflo (2001, 2004).[10] The large data set is especially important for our analysis because we follow Duflo (2001) to define the treatment (Inpres) and comparison (pre-Inpres) groups of children in a narrow window of age cohorts (5 years). Even with this cohorts restriction, our estimation sample consists of 2,048,164 father-child pairs. We provide evidence that the estimates from the census data do not suffer from any substantial bias relative to a widely used data set for estimating intergenerational effects in Indonesia: the Indonesia Family Life Survey (IFLS). IFLS includes information on nonresident household members, and thus is not subject to sample truncation.[11] We also report estimates using an inverse probability weighting (IPW) scheme that corrects for biases due to possible nonrandom sample truncation in census data even though such truncation bias seems small in our context.[12]

The evidence suggests five key conclusions. First, the conditional expectation function (henceforth CEF) for both the Inpres and pre-Inpres cohorts is concave when we estimate the influence of a father's education (years of schooling) on the probability of primary completion by children. The Inpres schools had a positive effect on the intercept, a negative effect on the linear term, and a positive effect on the quadratic term, and this sign pattern holds across gender. The primary completion estimates thus suggest that the children from the most disadvantaged background (fathers with no schooling) enjoyed higher relative mobility (negative effect on the linear term) and absolute mobility (positive effect on the intercept).[13] In our application, the standard linear model underestimates substantially the improvements in

[10]The influential contribution by Chetty et al. (2014) highlighted the advantages of big data in understanding intergenerational economic mobility. Card et al. (2022) use 1940 census data to analyze the effects of teacher quality on intergenerational educational mobility of black children in USA at the beginning of 20th century. A number of recent papers on intergenerational educational mobility in Africa use the 10 percent sample of the census available through IPUMS; see, for example, Alesina et al. (2021) and Azomahou and Yitbarek (2021).

[11]We do not use IFLS for our analysis because the sample size becomes too small once we restrict to the relevant cohorts. There are less than 250 household level observations available for some of our analysis where some districts have only 2 households in the sample.

[12]Nicoletti and Francesconi (2006) provide evidence that IPW performs better than Heckman selection correction in correcting coresidency bias in intergenerational mobility analysis.

[13]The intercept term represents the conditional expected years of schooling, a measure of absolute mobility, for children of fathers with no schooling. The linear term is the slope of the mobility CEF evaluated at father's zero schooling, and the slope of the mobility CEF is a measure of relative mobility.

relative mobility experienced by the most disadvantaged children: 72 percent underestimation for sons and 22 percent for daughters.[14]

Second, the estimates show important heterogeneity: while the Inpres schools improved relative mobility of the children from low educated households, it reduced relative mobility of children in the highly educated households. Since lower relative mobility implies a higher intergenerational persistence, Inpres schools strengthened the educational advantages of the more educated segment of the society across generations.

Third, in contrast to the primary completion results, there are dramatic gender differences when years of schooling is used as a measure of children's educational attainment. The CEF is concave for sons, but it is linear for daughters in this case. More importantly, the evidence suggests strong effects of Inpres schools on sons, but there are no significant effects on daughters.

Fourth, we explore alternative explanations for the puzzle of strong effects on primary schooling of girls, but little effects on their completed years of schooling. The 61,000 primary schools created a funnel effect because the number of primary graduates increased dramatically but the number of high schools did not expand in any significant way (Heneveld (1979)). Thus, the Inpres graduates faced an unintended bottleneck at the high school entry level. Our estimates of the effects of Inpres on high school completion suggest that the boys experienced a positive effect (7.7 percentage points higher probability of completion), but the girls suffered a negative effect (8.5 percentage points lower probability of completion). The expanded supply of primary graduate boys had crowded out the girls from the high school.[15]

Evidence suggests that the negative effect on a girl's secondary schooling does not depend on the level of her father's education. This implies that the mechanisms responsible for the gender-based crowding out are common to all children irrespective of family background. There are two such potential mechanisms: (i) low labor market returns for girls at the secondary and higher levels, and (ii) social norms against girls in a patrilineal society such as

[14] These estimates refer to primary completion of children. For years of schooling, the effect on the most disadvantaged sons is underestimated by 97 percent, but there is no significant effect on daughters in both linear and quadratic models.

[15] Once we take into account the differences in the base: for boys it is 0.197 and for girls 0.179, the positive effect for boys cancels out the negative effect on girls, suggesting a one for one gender based crowding out.

son preference in education.[16] A substantial body of evidence for the relevant cohorts in Indonesia shows higher returns to education for girls (e.g., Deolalikar (1993), and Behrman and Deolalikar (1995)), and thus rejects the labor market based explanation. This leaves gender norms as a plausible mechanism which we test using data from the "matrilineal island" West Sumatra, the home to the largest matrilineal tribe in the world (Minangkabau). If the crowding out observed at the national level is driven by gender bias against girls in the patrilineal islands, then we should not observe any crowding out of girls at the secondary level in the matrilineal island which is confirmed by the estimates.[17] Instead, the boys face significant crowding out in matrilineal West Sumatra.

Fifth, our estimating equations use years of schooling and primary completion, based on the recent quadratic mobility models developed by Becker et al. (2015), and Becker et al. (2018). We also provide evidence on alternative models of mobility, based on years of schooling normalized by its standard deviation and schooling ranks.[18] We find that the conclusions regarding the impact of Inpres schools on relative and absolute mobility of sons and daughters do not change substantially when we use the normalized model.[19] But the conclusions from the rank-rank mobility model are different, with no significant effects of school construction even for sons. We discuss how to interpret such conflicting evidence and provide guidance for advising policymakers.

The analysis and conclusions of this paper have wider implications. The evidence that the effects of government policies on relative mobility can be fundamentally different (with different signs) across educated and uneducated households is of more general interest. It underscores the importance of testing the default linear functional form of the mobility equation

[16]There is a growing recognition among economists that social norm is a first order factor for understanding the persistent biases faced by women in many developing countries (Jayachandran (2015)).

[17]We underscore here that we are concerned with gender norms for the 1950s to 1960s birth cohorts. Even though evidence suggests no significant gender bias in education in the recent decades in Indonesia (Afkar et al. (2020)), our findings about gender bias in these older cohorts are consistent with other available evidence. For example, Maccini and Yang (2009) find that girl's health and schooling outcomes in these cohorts were sensitive to early life rainfall shocks, but parents effectively insured the boys against such shocks, suggesting strong son preference.

[18]Following the influential work on the rank-rank model for income mobility by Chetty eat al. (2014), many authors are increasingly adopting a rank-rank specification for intergenerational educational mobility (see, for example, Neidhofer et al. (2018), Hilger (2015), Asher et al. (2023)). This, however, ignores the fact that, unlike income, education is a discrete variable with limited support. For a discussion on the difficulties in adopting the rank-rank model for education, see Ahsan et al. (2022).

[19]Relative mobility is measured by Pearson correlation in the normalized model.

as it assumes away such heterogeneous effects of a policy on relative mobility. The insight that the schooling expansion at the primary level may create an unintended bottleneck at the secondary level which in turn can lead to adverse distributional effects on historically disadvantaged groups is relevant for policymakers in many other countries. Because primary school expansion from 1970s onward has been dramatic in most of the developing countries but the expansion at the secondary and higher levels has lagged far behind.[20] The more general lesson here is that expanding educational opportunity at a lower level without concurrent targeted policies at the next level may result in crowding out of the children from disadvantaged background (e.g., ethnic and religious minority, lower caste, and girls). The potential conflict between rank-based mobility model and the models based on Becker-Tomes (years of schooling) is likely to be important for the evaluation of other government policies, and in many other countries.

The remainder of the paper is organized as follows. Section (2) discusses the related literature on intergenerational mobility and the effects of better access to schools to put the paper in perspective. Section (3) contains a description of the census 2000 data and the variables in our analysis. Section (4) lays out the empirical strategy and the estimating equations for the linear and quadratic mobility models. Section (5) reports the main estimates of the effects of Inpres schools on the mobility equations for sons and daughters, including estimates of the effects of Inpres schools on relative and absolute mobility. This section also offers robustness checks using different comparison groups, and mother's schooling in place of father's schooling as an indicator of children's family background. Section (6) reports the estimated effects on relative and absolute mobility of sons and daughters. Section (7) is devoted to uncovering the mechanisms behind the puzzling absence of any effect on girl's completed years of schooling notwithstanding the strong effects found at the primary level. This section also provides evidence on the sources of gender-based crowding out observed at the secondary level. Section (8) provides evidence using alternative models of mobility, based on schooling ranks and years of schooling normalized by generation specific standard deviation. This section also discusses how to advise policymakers when different models give conflicting evidence. The paper ends with a summary of the main findings and their implications for the

[20] In 2016, the secondary completion rate was only 35 percent in the low income countries as classified by World Bank. The corresponding rate for OECD countries was 96 percent. See chapter 2 in World Bank (2018).

broader literature on intergenerational mobility.

(2) Related Literature

The contributions of this paper are at the intersection of two major areas of economic research: intergenerational mobility and the effects of school expansion. There is a vast literature on intergenerational mobility in the context of developed countries, focusing primarily on intergenerational persistence in permanent income. See, for example, Solon (1992), Mazumder (2005), Chetty et al. (2014), Black et al. (2020), Carneiro et al. (2021), Acciari et al. (2022), Adermon et al. (2021), Abramitzky et al. (2021), Card et al. (2022), and Berman (2022). For excellent surveys, see Solon (1999), Black and Devereux (2011), Heckman and Mosso (2014), Mogstad and Torsvik (2021), and Cholli and Durlauf (2022). The literature on developing countries is limited, with most of the studies analyzing intergenerational persistence in educational attainment because of the paucity of long-run panel data required for a credible analysis of permanent income. Recent contributions in the context of developing countries include Agüero and Ramachandran (2020), Azam and Bhatt (2015), Alesina et al. (2021), Emran and Shilpi (2015), and Neidhofer et al. (2018).[21] Excellent surveys of the recent literature on developing countries are provided by Torche (2019), and Iversen et al. (2019).

A second strand of literature that our analysis contributes to is the effects of access to schools, especially public schools, on children's outcomes, including the intergenerational effects (for surveys, see Orazem and King (2008), Hanushek (2002), and Filmer (2007)).[22] Neilson and Zimmerman (2014) report that elementary and middle school construction projects in a poor urban school district in USA raised test scores, enrollment, and home prices. Currie and Moretti (2003) find that availability of college in a county in USA improved children's birth outcomes by increasing maternal education. Khanna (2023) analyzes the general equilibrium effects of a large scale school expansion program in India, and provides evidence that unskilled workers benefited while skilled workers were worse off.[23]

[21] Among unpublished papers, see Yu et al. (2020), Asher et al. (2023), Emran et al. (2021).

[22] *Intergenerational effects* refer to the effects of policies on the second generation: does improving education of parents by a policy improve the outcomes of their children? For an excellent discussion on the distinction between "intergenerational effects" and "intergenerational mobility", and the related literature, see Bjorklund and Jantti (2020).

[23] The sociological literature on education expansion (increase in average schooling) noted earlier focuses on intergenerational associations and "mechanistic" decomposition (not causal) to shed lights on the mechanisms. See, for example, Pfeffer and Hertel (2015) and Breen (2010). In the economic literature, education expansion

Following the influential work of Duflo (2001, 2004), many papers have studied the effects of the Inpres school construction in Indonesia; for example, Ashraf et al. (2020) find differential effects of school construction on daughters depending on whether dowry or bride price is practiced in the marriage market, Martinez-Bravo (2017) finds that school construction improved public goods provision, and Mazumder et al. (2019) and Akresh et al. (2018) analyze the "intergenerational effects", i.e., the effects on the educational and health outcomes of second generation (the children of the mothers who were exposed to Inpres schools as children). Both Mazumder et al. (2019) and Akresh et al. (2018) find substantial positive effects of higher mother's education who were exposed to Inpres program on the school performance and other outcomes of their children.[24] However, their focus is not on whether the the mothers exposed to Inpres schools were affected differently depending on their socioeconomic background. If the Inpres schools improved girls' education only at the more educated households, and this subsequently led to strong intergenerational multiplier effects on the second generation (as found by these papers), then the long-term effects of Inpres schools would be highly inequalizing. By focusing on how the advent of Inpres schools affected the link between education and family background for the children exposed to the program, we provide the critical missing link in understanding the long-term distributional consequences of the Inpres schools working through the intergenerational multiplier effect found in the earlier studies.

Assaad and Saleh (2018) provide evidence from Jordan that the influence of a parent's education on children's schooling is lower in a location where the number of basic schools (primary) is higher. But we cannot interpret their estimates as causal because, unlike Duflo (2001) (and our paper), there is no clear exogenous policy experiment that caused the school expansion in Jordan.[25] The school expansion was gradual, and is likely to reflect domestic economic conditions which might have affected the educational investments made by the parents (see the discussion in section (4) below). To the best of our knowledge, this is the first paper in the literature to study the effects of public school expansion on intergenerational educational mobility using a credible identification strategy and empirical specifications that

has been analyzed by Vella and Karmel (1999) for Australia and Blanden and Machin (2013) for United Kingdom.

[24] This positive multiplier effect is in contrast to Black et al. (2005) who does not find any causal impact of higher parental education on children's education in Norway except for the mother-son link.

[25] The authors acknowledge this limitation of their study.

allow for heterogeneous relative mobility across the distribution.

There is a rich literature on the effects of school quality and school reform on various outcomes of children. In a recent paper, Card et al. (2022) provide an interesting analysis of the effects of public schools in the early 20th century USA using 1940 census data. Their focus is on *quality of schooling* and they find that the impact of school quality varies by parental school level, supporting the emphasis on the heterogeneity in relative mobility across the distribution in our study. They provide causal evidence that salary caps reduced quality of teachers which affected the black children's educational attainment adversely. There are a few papers that look at the causal effects of school reform and/or school quality improvements on intergenerational income mobility using a difference-in-difference design; see Pekkarinen et al. (2009) in the context of a comprehensive schooling reform in Finland, and Parman (2011) for a historical study of Iowa showing that a better school quality lowered intergenerational income mobility.

(3) Data Description and Variables Definitions

The empirical analysis of this paper is based on Indonesian census 2000 full count data. Our main estimation sample consists of the children born between 1957 to 1962 and 1968 to 1972. The school construction under the SD Inpres program of the second five year plan began around 1973-1974. Following Duflo (2001), we define birth cohorts born between 1968 to 1972 as the exposed group as they are most likely to benefit from the program, and birth cohorts born between 1957 to 1962 as the comparison group as they are least likely to benefit from the program.[26] The intermediate birth cohorts 1963-1967 may be partially exposed to the Inpres schools, and we also provide estimates for this group. For our main analysis, we rely on father's schooling as a measure of parental education. As part of robustness checks, we also report estimates using mother's schooling as a measure of parental education. We underscore here again that parental schooling (education) in our analysis is a summary measure of the socioeconomic status of the family a child is born into, and thus capture any and all correlated family and neighborhood factors that affect children's educational attainment.

Given that we have the full count census data, our main estimation sample (treatment

[26] A child born in 1962 is 11-12 years old in 1973 and already completed primary schooling before the Inpres program.

group: 1968-1972 and comparison group: 1957-1962) gives us 2,048,164 father-child pairs where household heads are the fathers, and the children were living at the household at the time of the census. We calculate the years of schooling based on the education level a respondent has completed. A comparison shows that the children exposed to Inpres schools and their fathers have more education than comparison cohorts and their fathers (see Table 1).

The census also reports information on birth district and province of an individual. We match this birth district information with the Inpres school construction intensity data, which was graciously provided to us by Esther Duflo. The school intensity data was originally reported as number of schools per 1000 children in a district. We use a normalized measure of the Inpres intensity by dividing the number of schools by the highest number of schools received by a district.[27] This normalization implies that the estimated coefficients can be interpreted directly as the effects for the districts with the highest treatment intensity.

We use built-up density in a district in 1975 for estimating the selection equation for father and children's coresidency. We implement inverse probability weighting to correct the biases in the estimates due to sample truncation arising from coresidency. The source of built-up data is the Global Human Settlement Layer (GHSL) (Pesaresia et al. (2015)). The built-up data are at 300 meters by 300 meters grids. We super-impose the digital maps from the censuses on the pixel-level data to estimate the total built-up area at the district level.

(4) Empirical Strategy and Estimating Equations

Our empirical strategy follows closely the approach due to Duflo (2001, 2004) that exploits both cross sectional variation across districts and over time variation across birth cohorts. The cross sectional variation comes from the differences in the intensity of exposure depending on the number of new schools constructed in a district under the Inpres program. The allocation rule decided the number of new schools in proportion to the number of children of appropriate age group not enrolled in primary school in 1971 (Aziz (1990)). The over time variation comes from comparing birth cohorts that were exposed to the new schools with those who completed

[27] The district that received the highest number of schools received 8.6 schools per 1000 children. In contrast, the district that received the lowest number of schools received only 0.59 schools per 1000 children. The mean is 1.86 schools per 1000 children. After normalization, the school construction intensity value ranges from 0.0678 to 1 with a mean of 0.215.

schooling before the construction of the Inpres schools. A major concern here is whether the timing of the program implementation can be treated as exogenous. If the school constructions were undertaken by the government in response to some shocks to the domestic economy with differential effects across districts, then the same shock could affect the educational outcomes of children through family income independent of the effects of Inpres schools. The public funds for such a massive school construction program were generated by an external shock in the international market for crude petroleum (gasoline). The dramatic increase in the oil prices owing to the 1973 OPEC oil shock created a huge windfall for government of Indonesia, and the size of the Indonesian government budget increased 2.5 times from 1973 to 1975. The Inpres school construction under the second five year plan was thus not related to any domestic economic factors.

Following the influential work of Duflo (2001, 2004), the effects of Inpres school expansion have been studied by many papers, and we have a wealth of accumulated evidence on the validity of the quasi-experimental design originally developed by Duflo (2001).[28] The evidence reinforces and enhances the credibility of the research design in a variety of contexts using different data sets. Some of this evidence is directly relevant for the validity of the research design in our application. In the context of our analysis, an important issue is whether the high Inpres intensity districts were experiencing higher growth in educational outcomes in the pre-program period because of factors unrelated to school construction. Mazumder et al. (2019) use 1985 intercensal population survey (SUPAS 1985) data and show that there are no significant differences in the trend of primary completion rates across districts for the cohorts that completed primary schooling before Inpres school construction. This allays the concern of differential underlying trends across districts. In addition, we find that treatment intensity is not correlated with father's education which is informative about whether the program was systematically targeted to areas with low educated parents.[29] Another concern relates to concurrent government programs that might have affected children's educational outcomes and might be spatially correlated with Inpres intensity across districts. Duflo (2001) carefully considers such threats to the identifying assumption and includes controls for a water and

[28]see, for example, Martinez-Bravo (2017), Mazumder et al. (2019), Jung et al. (2021), Ashraf et al. (2020), Akresh et al. (2018).

[29]The estimates are available upon request.

sanitation program implemented under Inpres. We thus include controls for the water and sanitation program exposure across districts; see below for details.

We provide evidence on the plausibility of the parallel trends assumption using event study design (please see the two immediately following subsections for details). In addition to parallel trend assumption, a DiD design requires a second identifying assumption: the no anticipation assumption. The "no anticipation" in our context implies that the schooling of the comparison cohorts cannot be affected by anticipatory actions of parents or children themselves before the actual school opening in a village. However, it is not likely to be a concern in our application for the following reason. While it is possible that parents may invest time and money (say on books) in anticipation of a primary school opening in the village next year, such investments will be for the children in primary schooling age, not for the cohorts who completed primary schooling and constitute our comparison group. Such parental investments thus do no affect the final schooling attainment of the comparison cohorts.

(4.1) Estimating Equation: Linear Model

In this section, we rely on the linear intergenerational educational mobility model. The DiD empirical model in our application can be written as:

$$\begin{aligned} E^c_{ikt} =& \beta_0 + \beta_1 E^p_{ikt} + \beta_2 Exp_t + \beta_3 Inp_k + \beta_4 \left(Exp_t \times Inp_k\right) + \beta_5 \left(E^p_{ikt} \times Inp_k\right) \\ & + \beta_6 \left(E^p_{ikt} \times Exp_t\right) + \beta_7 \left(E^p_{ikt} \times Exp_t \times Inp_k\right) + \varepsilon_{ikt} \end{aligned} \tag{1}$$

where E_{ikt} is an indicator of educational attainment of child i and his/her parents (depending on the superscript), superscripts c and p refer to children and parents respectively, k denotes the birth district of child i, t denotes time period (year), Exp_t is a dummy that takes on the value 1 for the children exposed to the Inpres schools (born between 1968-1972) and zero otherwise (born between 1957-1962). Inp_k is a measure of the intensity of the new school construction in district k. We use a normalized measure so that $Inp_k \in [0, 1]$.[30] For details of the construction of the program intensity variable, please see section (3) above. For sons we use $c = s$, and for daughters $c = d$ as the superscript. The intercept effect of the Inpres schools is captured by β_4 and the slope effect by β_7.

[30]We note that the main conclusions of the paper do not depend on this normalization.

As discussed above, our identification strategy closely follows that of Duflo (2001, 2004). Following Duflo (2001, 2004), we include birth district fixed effects (α_k) and birth year fixed effects (τ_t) for a child, and the following interactions (denoted by $W_k \times \tau_t$): year of birth interacted with 1971 enrollment (before Inpres program), year of birth interacted with the number of school-age children in 1971, year of birth interacted with water sanitation program, with all the variables measured at the birth district level. Note that the district fixed effects absorb the level effect of the Inpres school intensity variable Inp_k as it varies at the district level, and the birth year dummies absorb the level effect of the Exp_t dummy. The estimating equation for the linear model becomes:

$$E^c_{ikt} = \beta_0 + \beta_1 E^p_{ikt} + \beta_4 \left(Exp_t \times Inp_k\right) + \beta_5 \left(E^p_{ikt} \times Inp_k\right) + \beta_6 \left(E^p_{ikt} \times Exp_t\right) + \beta_7 \left(E^p_{ikt} \times Exp_t \times Inp_k\right) + \alpha_k + \tau_t + \sum_t W_k \times \tau_t + \varepsilon_{ikt} \quad (2)$$

A large literature on intergenerational educational mobility focuses on the parameter β_1, called intergenerational regression coefficient (IGRC, for short) which is a measure of relative mobility (see, for example, Azam and Bhatt (2015), Neidhofer et al. (2018)). Our focus is on the parameter β_7 that captures the effects of Inpres schools on IGRC. Note that IGRC provides an estimate of intergenerational persistence in education, and, a higher persistence implies lower relative mobility. When $\beta_7 < 0$, Inpres schools improve relative mobility, the IGRC estimate for the exposed cohorts is smaller in this case, i.e., ($\beta_1 + \beta_7 < \beta_1$).

A focus of our analysis is on the effects of Inpres schools on the intercept of the mobility equation as captured by β_4. As noted earlier a higher intercept ($\beta_4 > 0$) implies a positive impact of the Inpres schools on the expected educational attainment level of children from the most disadvantaged households (fathers with no schooling). The intercept can be interpreted as a measure of absolute mobility of these disadvantaged children.[31]

We report evidence consistent with the parallel trends assumption for the linear mobility model in equation (2). Event study graphs for the two parameters of interest β_4 (intercept effect) and β_7 (slope effect) are in Figure 1 (sons) and Figure 2 (daughters). The event study

[31] Measuring absolute mobility by the expected outcomes of children based on a conditional expectation function (CEF) has been popularized by the recent work on Chetty et al. (2014) on intergenerational income mobility in USA. For details, see the discussion in section (5.3) below.

graph for sons show that the placebo effect of a fictitious school construction on the probability of primary completion of children at various years before the actual construction of schools is zero for both parameters of interest. In contrast, there is an appreciable shift in the estimated effects for the post school construction years for the fully exposed birth cohorts (1968-1972). The evidence is similar for girls.[32] In the online appendix, we report the event study graphs including the partially exposed cohorts (see Figures A.1 (sons) and A.2 (daughters)).

(4.2) Estimating Equation: Quadratic Model

Many existing studies on intergenerational educational mobility, both in economics and sociology, rely on the linear model discussed above (see, for example, Hertz et al. (2008), Lou and Li (2022), Azam and Bhatt (2015), Emran and Shilpi (2015)). For a survey, see Torche (2019)). However, recent theoretical and empirical analyses suggests that the linear model may be inadequate for understanding intergenerational educational mobility (Becker et al. (2015), Emran et al. (2021)). For recent contributions where mobility curve is nonlinear, and relative mobility and the effects of policy interventions vary across the distribution of father's education, see Card et al. (2022), Asher et al. (2023), Emran et al. (2021), and Ahsan et al. (2021).

Many authors use the estimate of relative mobility from a linear model as a summary measure because it can be interpreted as an weighted average of underlying heterogeneous relative mobility of different subgroups across the distribution. However, recent work by Maasoumi et al. (2022) shows that the weights implied by the OLS estimate of the linear mobility equation have no plausible economic interpretation. Our objective here is not to provide a summary measure, but to understand potentially very different effects of policies on different sub groups defined by father's schooling level.

To the best of our knowledge, the quadratic intergenerational educational mobility model was first derived by Becker et al. (2015). They allow for an interaction effect in the education production function where marginal returns to financial investment in children's education increase with the level of parental education. This complementarity between parental edu-

[32] The event study graphs for years of schooling also show that the placebo treatment effects for the pre-intervention birth cohorts are not significantly different from zero at the 10 percent level. We also checked the placebo effects for the earlier birth cohorts in the pre-Inpres period (1950-1956), and the results support the DiD design. The details are available from the authors.

cation and financial investment can arise from a variety of sources and make the mobility curve convex. The sources of complementarity include role model and peer effects, and more efficient educational investments by educated parents. The returns to financial investment for a given level of father's education however are subject to diminishing returns which can result in a concave mobility CEF when the forces of complementarity are weak or nonexistent. As noted earlier, for our analysis, an important question is whether the advent of Inpres schools changed the shape and degree of nonlinearity in the mobility curve, because such a change may result in very different effects on the children of low educated parents relative to the effects on the children born to highly educated parents.

For the quadratic intergenerational educational mobility model, the DiD empirical specification is (with fixed effects):

$$
\begin{aligned}
E^c_{ikt} = {} & \theta_0 + \theta_1 E^p_{ikt} + \theta_4 \left(Exp_t \times Inp_k\right) + \theta_5 \left(E^p_{ikt} \times Inp_k\right) + \theta_6 \left(E^p_{idt} \times Exp_t\right) + \theta_7 \left(E^p_i \times Exp_t \times Inp_k\right) \\
& + \theta_8 \left(E^p_{ikt}\right)^2 + \theta_9 \left(\left(E^p_{ikt}\right)^2 \times Inp_k\right) + \theta_{10} \left(\left(E^p_{ikt}\right)^2 \times Exp_t\right) + \theta_{11} \left(\left(E^p_{ikt}\right)^2 \times Exp_t \times Inp_k\right) \\
& + \alpha_k + \tau_t + \sum_t W_k \times \tau_t + \zeta_{ikt} \qquad (3)
\end{aligned}
$$

The focus here is on three parameters: θ_4 (the intercept effect), θ_7 (effect on the linear term), and θ_{11} (effect on the quadratic term). In a quadratic mobility model, the impact on the constant provides an estimate for the effects of Inpres schools on absolute mobility of the children of fathers with no schooling, similar to the linear model. The main difference from the linear model is that relative mobility varies across the distribution of a father's schooling. The estimate of θ_7 is the effects on relative mobility of the children of fathers with no schooling, but relative mobility of children of fathers with positive schooling depends on both the linear and quadratic effects: θ_7, and θ_{11}. This allows for the possibility that the effects of Inpres schools on relative mobility can have opposite signs at the tails of the distribution.

We provide evidence supporting the parallel trends assumption in the context of the quadratic mobility model in equation (3). We report event study graphs for the three parameters of interest β_4, β_7 and β_{11}. The event study graphs for sons (Figure 3) show that the null hypothesis of no effect of fictitious school construction at various years before the actual construction of schools cannot be rejected at the 5 percent level for the parameters of interest.[33] In contrast, there is a clear shift in the estimated effects for the post school construction years, and the effects are significant at the 5 percent level for all three parameters. The event study graphs for daughters (Figure 4) suggest a significant shift in the intercept after the school construction, but the effects on the linear and quadratic coefficients seem much weaker.[34] This alerts us about potentially important gender differences in the effects of school construction. Even study graphs including the partially exposed cohorts are reported in the online appendix (see Figures A.3 (sons) and A.4 (daughters) in the online appendix).

(5) Empirical Evidence: Estimates of the Effects of Inpres Schools on the Mobility CEF

We report the estimates of equations (2) and (3) with two alternative measures of children's educational attainment: years of schooling and a binary indicator for primary or more schooling. Completed years of schooling is the most widely used measure of educational attainment of children in the literature on intergenerational mobility. As noted earlier, one can argue that the most relevant indicator to judge the effects of the new primary schools is whether a child completed primary schooling. As a measure of parental education, we rely on the completed years of schooling of the father of a child in all our analysis, keeping with a large literature on intergenerational educational mobility (see the surveys by Torche (2019), Iversen et al. (2019), and Emran and Shilpi (2021)). All reported standard errors are clustered at the district level, following Duflo (2001).

(5.1) The Effects of Inpres Schools on Years of Schooling of Children

Table 2 reports the estimates of the mobility equation when children's educational attainment is measured by their completed years of schooling which is standard in many studies of intergenerational educational mobility (see the survey by Torche (2019)). The first two columns in Table 2 report the estimates from the linear model, while the last two from the quadratic model. We report the estimated effects of Inpres schools on the parameters of interest (intercept, linear, and quadratic terms of the mobility models), and the full set of

[33] The placebo effects are not significant at the 10 percent level also.

[34] The corresponding event study graphs for years of schooling also support the plausibility of the DiD design. We also checked the event study graphs including earlier birth cohorts in the pre-Inpres period (1950-1956) and the evidence is consistent with the graphs reported in the online appendix. The details are available from the authors.

coefficient estimates for the corresponding DiD models are reported in online appendix Table A.1.

The estimates from the linear model suggest that, for sons (superscript s denoting sons), there is a positive effect on the intercept $\left(\hat{\beta}_4^s = 1.59\right)$, and a negative impact on the slope (IGRC) $\left(\hat{\beta}_7^s = -0.13\right)$.[35] The evidence on the slope thus suggests that the Inpres schools weakened the impact of family background and improved relative mobility of boys irrespective of a father's education level. The positive effect on the intercept suggests that the most disadvantaged children (fathers with no schooling) benefited in the form of higher expected years of schooling as a result of the Inpres schools. However, the estimated effects for daughters are numerically much smaller and are not significant at the 10 percent level, suggesting that the expansion of the primary schools failed to affect educational mobility of the girls in a significant way. This absence of a significant effect for girls is unexpected because a substantial literature suggests that availability of schools in a village is more important for girls.

Linearity is a maintained assumption in the estimates in columns (1) and (2) of Table 2, but a growing theoretical and empirical literature suggests that the intergenerational education mobility CEF may be concave or convex. Once we admit the possibility that the mobility CEF could be concave or convex, an important question is whether the school constructions affected the degree of concavity (or convexity) and whether there is any gender differences in the changes in the shape of the CEF. Table A.2 (panel A) in the online appendix reports estimates of a standard quadratic mobility model (see, for example, equation (8) of Becker et al. (2015)) for the pre- and post cohorts in our data to understand whether a linear CEF is a reasonable approximation for evaluating the effects of Inpres schools on children's completed years of schooling. For daughters, the evidence suggests that the CEF is linear for both the pre-Inpres and Inpres cohorts, thus suggesting that the linear DiD model in column (2) of Table 2 is appropriate for the analysis of daughters. The evidence for sons is different: the mobility CEF for sons was approximately linear in the pre-Inpres cohorts, but it has become significantly convex in the Inpres cohorts. So we need to allow for a quadratic specification for sons.

Column (3) in Table 2 reports the estimates of the main parameters of interest from the

[35] The intercept effect is significant at the 1 percent level, and the slope effect at the 5 percent level.

quadratic model (estimating equation (3)) for sons and the estimated full specification is provided in the online appendix Table A.1. The evidence form the quadratic model suggests a positive effect on the intercept $\left(\hat{\theta}_4^s = 1.53\right)$ which is numerically close to the estimate from the linear model (significant at the 1 percent level). The estimated impact on the linear coefficient is negative and 100 percent larger in magnitude compared to the linear model $\left(\hat{\theta}_7^s = -0.26\right)$ (significant at the 1 percent level). *This indicates that the linear model substantially underestimates the improvements in the relative mobility of the most disadvantaged boys born to fathers with no schooling.*[36] The evidence suggests that the Inpres schools had a positive effect on the quadratic coefficient making the CEF convex $\left(\hat{\theta}_{11}^s = 0.015\right)$ (significant at the 5 percent level). As we will see in section (5.3) below, a positive effect on the quadratic coefficient implies that the effects of Inpres schools on relative mobility are opposite for the sons in uneducated households vs. the sons in highly educated households.

(5.2) The Effects on Inpres Schools on the Completion of Primary Schooling

As noted earlier, a natural metric to measure the effectiveness of the new primary schools is to look at the primary schooling completion of the exposed cohorts of children. The estimates of the parameters of interest in equations (2) (linear model) and (3) (quadratic model) are reported in Table 3 (the full set of coefficients are reported in online appendix Table A.3).

We first consider the estimates from the linear model. For sons, the pattern of the effects on the probability of having primary or more schooling are similar to what we found earlier using years of schooling as a measure of educational attainment. But the evidence is dramatically different for daughters: there are numerically substantial and statistically significant (at the 5 percent or less) effects on both the intercept and the slope (IGRC).[37] However, these results are built on the maintained assumption of a linear CEF which we test next.

To determine whether the linear model is appropriate, we estimate a standard quadratic mobility model for the pre-Inpres and Inpres cohorts separately using a dummy for primary or more schooling as the measure of children's educational attainment. The estimates are reported in the online appendix (see panel B of Table A.2). The evidence suggests that the

[36]Recall that the linear coefficient in a quadratic model gives the IGMA estimate for the children of fathers with no schooling.

[37]The slope estimate from a linear educational mobility is called intergenerational regression coefficient (IGRC for short) in the literature.

mobility CEF is concave irrespective of gender for both the pre-Inpres and Inpres cohorts, and, perhaps more interesting, the degree of concavity has declined after the school construction. This indicates that the estimates from the linear model may be seriously misleading in this case, and a quadratic model would be more appropriate for both sons and daughters to understand the effects of Inpres schools on their primary school completion.

The estimates of the relevant parameters from the quadratic model are reported in columns (3) (sons) and (4) (daughters) of Table 3. The pattern of the estimated effects of Inpres are similar across gender: positive for the intercept $\left(\hat{\theta}_4^s = 0.195;\ \hat{\theta}_4^d = 0.197\right)$, negative for the linear coefficient $\left(\hat{\theta}_7^s = -0.038; \hat{\theta}_7^d = -0.017\right)$, and positive for the quadratic coefficient $\left(\hat{\theta}_{11}^s = 0.002; \hat{\theta}_{11}^d = 0.001\right)$. The effect on the curvature of the mobility equation for daughters is much weaker: the estimated effects on the linear and quadratic coefficients are about half of the corresponding estimates for sons. For sons, all three coefficients are significant at the 1 percent level, while, for daughters, the intercept effect is significant at the 1 percent level and the linear and quadratic effects at the 10 percent level. This suggests that the effects of Inpres schools vary substantially across family background for sons, but such heterogeneity is much less important for daughters which is consistent with the idea that girls schooling is largely determined by social norms regarding gender roles. We explore the role of social norms in more depth and details later in the paper.

The estimated quadratic coefficients look small in magnitude, especially compared to the linear coefficients, and a reader might wonder whether the linear model is after all a good approximation for the evaluation of the effects of Inpres schools. However, note that the impact on relative mobility due to the quadratic coefficient equals $2\hat{\theta}_{11}^c E_i^c$. This implies that the impact for a son whose father has 9 years of schooling is 0.038 which equals the the linear coefficient in magnitude $\left(\hat{\theta}_7^s = -0.038\right)$. Similar conclusions hold for the daughters' estimates. We provide estimates of changes in relative and absolute mobility due to the advent of the Inpres schools in section (6) below.

(5.3) Robustness Checks and Other Concerns

We check the robustness of our estimates above in a number of ways. First, we use mother's schooling instead of father's schooling as an indicator of children's socioeconomic status. This is motivated by substantial evidence that mother's influence is stronger on daughters (Torche

(2019), Emran and Shilpi (2011)). The evidence, however, suggests that the results and the conclusions do not change in any significant manner. Please see subsection (OB.1.1) in the online appendix.

The second robustness check uses alternative comparison groups, by excluding the oldest birth cohorts who might be less comparable. Again, the conclusions are robust. Please see online appendix subsection (OB.1.2) for the details.

The third robustness check addresses the issue of potential biases in the estimates from nonrandom sample truncation owing to coresidency in the census data. We provide evidence on this issue from two approaches, and the evidence taken together suggests that our conclusions are unlikely to be driven by coresidency bias. First, we use rich data from Indonesia Family Life Survey (IFLS) which includes information on nonresident parents and children, and thus do not suffer from cooresidency bias. A comparison of the census estimates of the mobility equations with those from IFLS shows that the estimates are in general close, and the degree of downward bias in the census estimates is small. Second, we implement correction for possible truncation bias (even if small) by using inverse probability weighting (IPW). Nicoletti and Francesconi (2006) provide evidence that IPW performs better than Heckman selection correction when dealing with such coresidency bias in intergenerational mobility analysis. They suggest house rental cost for identification, as house rental is the largest cost for children planning to leave parental home. Since house rental rates are not available for our study period, we use the recently available data on built-up density. Built-up density in 1975 in a district is used as a measure of house rent: a higher built-up density in a district implies lower rent and a lower coresidency rate (borne out by a negative coefficient on built-up density in the selection equation). The IPW corrected estimates are similar to the unweighted estimates reported above. We provide an explanation for the evidence that selection correction does not alter the estimates substantially. This apparently surprising finding can be understood better when we look at the correlation between Inpres school intensity and coresidency rates across districts: the inpres schools had no significant effect on the coresideny rates. For a detailed discussion and the relevant evidence, please see online appendix section (OB.2).

(6) The Effects of Inpres Schools on Relative and Absolute Mobility

In this section, we discuss the effects of Inpres schools on relative and absolute mobility. We focus on the estimates for primary or more education of children as the relevant measure of educational attainment, as it is the most natural metric to judge the effectiveness of a primary school. The evidence that the Inpres primary schools reduced the linear coefficient of the intergenerational educational mobility equation but increased the quadratic coefficient suggests that the effects on relative mobility could be very different at the two tails of father's schooling distribution.

For years of schooling, the effects on relative and absolute mobility of sons are similar to those found for primary completion and are not discussed here for the sake of brevity. Please see online appendix Table A.4 for the effects on sons' relative and absolute mobility when years of schooling is the measure of educational attainment. For daughters, there are no significant effects on relative or absolute mobility in terms of completed years of schooling because the estimates are not significant at the 10 percent level and much smaller in magnitude (Table 2).

(6.1) Effects on Relative Mobility

Since the mobility CEFs are concave for both sons and daughters for primary completion, we extend the standard measure of relative mobility in a linear model called intergenerational regression coefficient (IGRC). In a quadratic mobility model, a natural extension of IGRC is intergenerational marginal association (IGMA, for short) which is the slope of the mobility CEF at each level of father's education (see Emran et al. (2021)). This follows a large literature in economics and sociology where relative mobility is measured by the slope of the CEF relating children's economic status to that of their parents.[38] In a district with Inpres intensity of 1, relative mobility of the children of fathers with y years of schooling is given as (based on equation (3)):

$$IGMA_y = \theta_1 + 2\theta_8 E^p_{iy} \qquad PreInpres\ Cohorts$$
$$IGMA_y = \theta_1 + \theta_7 + 2\left(\theta_8 + \theta_{11}\right) E^p_{iy} \qquad Inpres\ Cohorts$$

[38] See Solon (1999) for the economic literature and Torche (2015) for the sociology literature. The most widely used measure of relative mobility in the literature on intergenerational income mobility is intergenerational elasticity (IGE) which is estimated as the slope of a log-linear CEF. The recent influential work of Chetty et al. (2014) estimates relative income mobility as the slope of a rank-rank CEF.

Using the estimated coefficients for primary completion in Table 3, we calculate the change in intergenerational marginal association, IGMA, for children of different socioeconomic background as represented by the level of father's schooling. The change in IGMA for the children of fathers with y years of schooling because of Inpres schools at treatment intensity 1 (the highest intensity) is give by:

$$\Delta IGMA_{iy} = \theta_7 + 2\theta_{11} E_{iy}^p$$

We provide estimates of change in IGMA for two levels of treatment intensity: the mean level and the highest intensity in our data.[39]

The estimates of the effects of Inpres schools on relative mobility are reported in Table 4. The evidence suggests that the Inpres schools increased relative mobility (lowered the IGMA) of the children from low educated households. The effect on the IGMA is the largest for the children born into the most disadvantaged households with fathers having no schooling. In this subgroup, the IGMA_0^s (subscript denoting the schooling level of fathers) for a son growing up in a district of the highest Inpres intensity declined by 3.8 percentage points (see row 1 and column 1), and by 0.80 percentage points in a district of average intensity (see row 1 and column 3). The corresponding estimates for daughters are smaller: 1.7 percentage points for the highest treatment intensity and 0.40 percentage points for the average intensity.

Without comparing to a benchmark, it is not clear whether these are substantial effects. We use the IGMA of the children of average socioeconomic background (fathers with average schooling) in the districts without any Inpres schools (zero Inpres intensity) in the pre-Inpres period as the benchmark.[40] The normalized estimates relative to the benchmark are reported in the even numbered columns in Table 4. For the sons born to fathers with no schooling, the Inpres schools reduced the IGMA by 33.96 percent in a district with mean Inpres intensity (row 1 and column 4), and by 157.95 percent in a district with the highest treatment intensity (row 1 and column 2). For daughters in this subgroup (fathers with no schooling), the corresponding improvements in relative mobility (reductions in IGMA_0^d) are much smaller in magnitude:

[39] The highest (normalized) intensity in our data is 1. The mean is 0.22 which correspond to 1.86 new schools per 1000 school age children in a district.

[40] The corresponding estimates using an alternative benchmark, the comparison (unexposed) children *born to fathers with no schooling* in districts with zero Inpres intensity, are similar. The estimates are available upon request.

15.19 percent (average intensity; see column 8, row 1) and 70.66 percent (highest intensity; see column 6, row 1).

In contrast, the relative mobility of the children of college educated fathers (16 years) worsened as a result of the Inpres schools and the role played by father's education increased: the IGMA_{16}^{s} for sons is 26.81 percent higher in a district with average treatment intensity, and 124.70 percent higher in a district with the highest Inpres intensity (see first four columns of row 5). Similar conclusions hold for daughters even though the magnitudes are smaller (see the last four columns of row 5). This suggests that the inheritance of educational status became more persistent across generations at the top of the education distribution, especially for the sons.

(6.2) Effects on Absolute Mobility

Absolute mobility is measured by the expected educational attainment of children conditional on father's education which is given by the point on the estimated CEF corresponding to a given level of father's education. This measure follows the recent influential work of Chetty et al. (2014).[41]

The estimated effects of Inpres schools on absolute mobility with primary completion as a measure of children's educational attainment are reported in Table 5 for two different levels of treatment intensity (average and the highest). To understand the magnitudes of the effects, as the base, we use the expected educational outcome of the children with average socioeconomic background (fathers with mean schooling) in the pre-Inpres cohorts living in the districts with no Inpres schools.

The estimates suggest that the new schools improved substantially the expected educational outcomes of children at two tails of father's education distribution but the effect was small at the middle of the distribution. For the children from most disadvantaged background (fathers having no schooling), there is no gender gap (see row 1 of Table 5). In this most disadvantaged group, both the sons and daughters experienced a 21 percent higher probability of having primary or more schooling in districts with the highest treatment intensity, and about 4.5 percent higher probability at the districts with average Inpres intensity (relative to

[41] Their P25 measure of absolute mobility is the expected income rank of the children conditional on father being in the 25th percentile of the income distribution.

the expected years of schooling of the benchmark group). In contrast, there are clear gender differences in the college educated households: the daughters benefited 60 percent more than the sons (at the highest Inpres intensity): the daughters reaping a 21.34 percent higher probability of having primary or more education while the sons gaining about 13.38 percent higher probability.

(7) The Consequences of an Unintended Bottleneck: Understanding the Sources of Gender Differences

A striking finding from our analysis above is that the effects of Inpres schools are dramatically different for daughters across primary vs. final educational attainment (completed years of schooling). In contrast, the effects are broadly similar for sons. The goal of this section is to understand what mechanisms can give rise to this gender difference.

The construction of 61,000 new primary schools increased substantially the supply of students competing for entry into high schools, but there were no significant expansions in the availability of high schools in Indonesia during the relevant period (Heneveld (1979)). This created an unintended bottleneck at the secondary schooling level. This raises a natural question: what were the effects of Inpres primary schools on educational opportunities beyond the primary level? A plausible conjecture is that the role played by socioeconomic background might have increased in the face of higher competition for a limited number of high school slots. If this is the case, we will see the children from more educated households gaining at the expense of low educated households, as the more educated parents usually have higher income and a more effective and extensive social network.[42] A testable implication of this hypothesis is that the impact of father's schooling at the high school level should become much stronger with the advent of an Inpres school in a district, irrespective of gender. However, this mechanism cannot lead to gender differences.

A second hypothesis focuses on gender specific constraints. As noted earlier, there is a substantial literature suggesting that distance to schools matters much more for schooling decisions of girls.[43] If cost of safe transport to far away schools is the binding constraint, then we would still expect substantial effects of family background, where only the girls from low

[42] One would expect bribery and "donations" to play important role in who gets admitted in the case of such a bottleneck.

[43] See the discussion by Scott (1985) in the context of Indonesia in the 1970s and 1980s.

educated (and low income) households fail to progress beyond primary schooling because they cannot afford safe transport for the daughters.

To explore these questions, we estimate the effects of Inpres schools on high school completion (12 years or more schooling) by children.[44] The estimates from the quadratic model are reported in Table 6.

The full sample (national) results in Panel A of Table 6 suggest a striking finding: family background (as measured by father's education) plays virtually no role in determining the impact of Inpres school expansion on higher secondary schooling: the triple interactions of Inpres intensity with both father's education and father's education squared are not significant at the 10 percent level and small in magnitude (compared to the estimates for primary schooling). These conclusions hold irrespective of the gender of a child. The estimated effects of Inpres on the intercepts suggest a different picture: it is negative and statistically significant for girls. The negative effect is substantial in magnitude: a 8.5 percentage points lower probability of having senior high schooling for the Inpres cohorts (panel A of Table 6).

The evidence for the sons in contrast suggests that they have gained: the probability of senior high schooling increased by 7.7 percentage points. The evidence thus indicates that the boys crowded out the girls at the senior secondary level. When we take into consideration that the base for boys (0.197) is larger that that for girls (0.179), the estimates suggest a one for one gender-based crowding out at the higher secondary level. The negative impact of Inpres schools on the girls at the secondary level explains the apparent puzzle of a substantial positive effect at the primary level and no significant effect on completed years of schooling.

(7.1) The Pattern of Crowding Out: Gender Bias in the Labor Market or Social Norm?

The evidence that the gender-based crowding out happened across the board suggests that it is primarily due to factors *which are gender specific, and unrelated to the socioeconomic background of a child*. There are two potential explanations. The first hypothesis is that in the 1970s and early 1980s, most of the parents considered primary completion as a social norm for girls, and paid little attention to their progression beyond this level. The

[44]Note that the linear mobility model is rejected for high school completion as a measure of educational attainment of children. Please see online appendix Table A.5 for details.

second explanation is based on possible gender differences in the labor market. If there were limited labor market opportunities and the returns to education were low for girls with secondary or more schooling in the 1980s, then we would expect the advent of Inpres schools to increase primary completion rate for girls, but still many of them would not go on to finish higher secondary schooling. There is substantial evidence that rejects this labor market based explanation. There was a considerable expansion of jobs for educated girls, especially in civil service; the proportion of women civil servants increased from 18% in 1974 to 27% in 1984 (see Calkins and Sengupta (1992)). Moreover, returns to education estimates for the relevant cohorts suggest *higher returns for women* at post primary levels in Indonesia (Deolalikar (1993), Behrman and Deolalikar (1995)).

The above discussion leaves us with social norm as a plausible explanation for the gender-based crowding out irrespective of family background. A credible way to test this explanation is to check if the gender-based crowding out is different between patrilineal and matrilineal tribes. If patrilineal social norm against women's higher education and participation in the formal labor market are driving the results we found in Table 2 earlier, then we should not observe any gender penalty against girls at the secondary level in a matrilineal society; if anything, we expect crowding out of sons in this case.[45] Indonesia is an excellent case study to test this hypothesis, as its West Sumatra province is the "Matrilineal island" where the largest matrilineal tribe in the world Minangkabau resides. We report the estimated impacts of Inpres schools in West Sumatra, and compare and contrast with the estimates from the other primarily patrilineal provinces separately in Tables 6 (panel B) for higher secondary completion. The estimates are strikingly different between the matrilineal vs. patrilineal islands. The crowding out of girls we observed at the national level seems to be driven solely by the patrilineal islands, while there is no negative impact for daughters in West Sumatra.[46] The sons in West Sumatra experienced a significant negative impact of Inpres at the secondary level, consistent with the expectation of a reversal of gender-based crowding out observed in the patrilineal islands (and at the national level).

[45]As noted in the introduction, we are concerned with gender norms for the 1950s- 1960s birth coborts in Indonesia. A substantial literature suggests that there are no significant gender differences in educational investments in the recent cohorts in Indonesia (see the discussion in Afkar et al. (2020)).

[46]This is consistent with Ashraf et al. (2020) which finds that Inpres schools had a much stronger positive effects on girls where bride price was practiced, i.e., the matrilineal tribes.

(8) Alternative Models of Mobility: Normalized Schooling and Ranks

Our analysis so far is based on years of schooling as a measure of children's final educational attainment. However, it is sometimes argued that years of schooling divided by generation specific standard deviation of schooling is preferable because the normalization makes cross-sectional inequality constant (equal to 1) across generations (Salvanes (2023), Torche (2019)). Some recent studies on intergenerational educational mobility adopt ranks in schooling distribution in each generation, following the influential work on intergenerational *income* mobility in USA by Chetty et al. (2014) (see, for example, Hilger (2015) on USA, Asher et al. (2023) on India, and Andrade and Thomsen (2018) on Denmark and USA). A recent paper by Ahsan et al. (2022) provides a comparative analysis of these alternative models of educational mobility with empirical evidence from China, Indonesia, and India. They show that the common argument that the rank-rank model removes the changing inequality across generations is valid for income but not for education. Percentile ranks in schooling distribution calculated by mid rank method often fails to equalize inequality across generations, because, unlike income, schooling is a discrete variable with limited support. They also show that the shape of the rank-based CEF may be fundamentally different from that based on years of schooling (normalized or not).[47]

To check whether the conclusions change substantially depending on the mobility model, we report the estimates for the rank-based and normalized models in the online appendix Table A.6. The estimates for the normalized model in the upper panel lead to the same conclusions we had earlier based on years of schooling: Inpres schools improved relative mobility of sons but had no significant effect on daughters. Even when the focus is on a measure of relative mobility that removes changes in cross-sectional inequality across generations, the evidence thus suggests that school construction improved relative mobility of sons and had no significant impacts on the daughters.

However, the rank-based model yields very different conclusions: there is no longer any significant effect on relative mobility irrespective of the gender of a child. This conflict between the rank-based model vs. the other two models may seem puzzling. But as discussed by Ahsan et al. (2022), the rank-based measure of relative mobility captures very different

[47] Ahsan et al. (2022) report evidence for India that the CEF for schooling ranks is convex while the CEFs are concave for years of schooling and normalized schooling.

mechanisms because it removes the economic forces affecting the marginal distributions (see also Torche (2013)). The rank-based measure primarily captures the "internal structure" of a society's educational opportunity, determined by formal and informal institutions. We expect the effects of school construction to work predominantly through the changes in the marginal distributions.[48] The evidence of a null effect suggests, perhaps not surprisingly, that school construction failed to alter the deep-seated institutional matrix relevant for educational inequality and mobility in Indonesia.

The conflicting evidence raises an uncomfortable question for a policy analyst: should we advise a policymaker that school construction was ineffective because it failed to affect the rank-based measure of relative mobility? When confronted with such conflicting evidence, policy advice can be grounded on the basic principle of the inequality of opportunity (IOP) as developed by Roemer (1998) (see also Coleman et al. (1966)): children should not be held responsible for the "circumstances" inherited by birth. In this perspective, a policy is effective if it weakens the impact of inherited circumstances such as father's education on children's educational opportunities. Since school construction in Indonesia weakened the impact of a father's education on his sons' schooling attainment, the policy should be considered effective, even though it failed to influence the deep-seated institutional matrix in Indonesia in the 1970s and 1980s.[49] The Inpres schools were effective for girls also in the sense that the girls born to uneducated fathers had a much higher probability of primary completion when exposed to the new schools.

(9) **Conclusions**

Exploiting a dramatic expansion of primary schools in Indonesia that doubled the number of primary schools in five years, we provide evidence on an important policy question relevant for most of the developing countries: does public investment in primary school construction improve intergenerational educational mobility of the disadvantaged groups such as girls and children born to uneducated parents? We take advantage of a large data set from the full

[48] Using a decomposition approach based on Shapely value, Apouey et al. (2022) provide evidence on 8 European countries that most of the cross-country differences in educational mobility are accounted for by the differences in marginal distributions.

[49] Note that schooling rank of a child can also be treated as an inherited circumstance even though the literature on IOP does not usually use ranks as circumstances. From this perspective, if a policy weakens the correlation in ranks between parents and children, this clearly implies improvements in mobility.

count census 2000 and rely on a credible identification scheme developed by Duflo (2001). Our empirical specifications are based on recent theoretical analysis that suggests that the intergenerational educational mobility curve can be concave or convex, with the workhorse linear model as a special case.

We find that Inpres schools made the mobility CEF less concave (or more convex), except for girls' years of schooling for which there is no significant effect on the shape of the mobility curve. This led to substantial improvements in relative mobility of the most disadvantaged children (fathers with no schooling) at the primary completion level irrespective of gender, but also reduced relative mobility at the more educated households. The new schools thus resulted in a stronger persistence in the educational advantages enjoyed by the most educated segment of the society.

The effects on the educational opportunities of children beyond the primary level are, however, dramatically different across gender. While the effects on sons' completed years of schooling are qualitatively similar to that on primary completion, surprisingly, there are no significant effects on girls. We explore the mechanisms behind the puzzling discord between the effects at the primary level versus competed years of schooling for girls. We find that the expansion at the primary level created a severe bottleneck at the secondary level, and the girls lost out facing fierce competition for a limited number of secondary school slots. The evidence suggests that the boys crowded out the girls at the senior secondary level, and the crowding out was experienced by girls irrespective of her father's schooling level. Contrasting evidence from the matrilineal island West Sumatra and the other patrilineal islands favors the hypothesis that the primary mechanism behind the gender-based crowing out is social norms against girls in a patrilineal society. The insight regarding unintended bottleneck and its perverse distributional consequences are of general relevance; any public policy that expands opportunity at a given level may end up unleashing economic forces that crowd out weaker social groups at the next level. To deal with such adverse crowding out, it is necessary to implement complementary policies targeting the social groups at risk of such crowding out.

References

Abramitzky, R., Boustan, L., Jácome, E., and Pérez, S. (2021). Intergenerational mobility of immigrants in the united states over two centuries. *American Economic Review*, 111(2):580–608.

Acciari, P., Polo, A., and Violante, G. L. (2022). And yet it moves: Intergenerational mobility in italy. *American Economic Journal: Applied Economics*, 14(3):118–63.

Adermon, A., Lindahl, M., and Palme, M. (2021). Dynastic Human Capital, Inequality, and Intergenerational Mobility. *American Economic Review*, 111(5):1523–1548.

Afkar, R., Yarrow, N., Surbakti, S., and Cooper, R. (2020). *Inclusion in Indonesia's education sector: A subnational review of gender gaps and children with disabilities*. The World Bank.

Agüero, J. M. and Ramachandran, M. (2020). The intergenerational transmission of schooling among the education-rationed. *Journal of Human Resources*, 55(2):504–538.

Ahsan, N., Emran, M. S., and Shilpi, F. (2021). Complementarities and Intergenerational Educational Mobility: Theory and Evidence from Indonesia. MPRA Paper 111125, University Library of Munich, Germany.

Ahsan, N. M., Emran, M. S., Jiang, H., Murphy, O., and Shilpi, F. J. (2022). When Measures Conflict: Towards a Better Understanding of Intergenerational Educational Mobility. Working paper, The World Bank.

Akresh, R., Halim, D., and Kleemans, M. (2018). Long-term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia. NBER Working Papers 25265, National Bureau of Economic Research, Inc.

Alesina, A., Hohmann, S., Michalopoulos, S., and Papaioannou, E. (2021). Intergenerational Mobility in Africa. *Econometrica*, 89(1):1–35.

Andrade, S. B. and Thomsen, J.-P. (2018). Intergenerational educational mobility in denmark and the united states. *Sociological Science*, 5:93–113.

Apouey, B., Nissanov, Z., and Silber, J. (2022). Ordinal variables and the measurement of upward and downward intergenerational educational mobility in european countries. *Review of Income and Wealth*.

Asher, S., Novosad, P., and Rafkin, C. (2023). Intergenerational Mobility in India: Estimates from New Methods and Administrative Data. *American Economic Journal: Applied*

Economics. Forthcoming.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2020). Bride price and female education. *Journal of Political Economy*, 128(2):591–641.

Assaad, R. and Saleh, M. (2018). Does Improved Local Supply of Schooling Enhance Intergenerational Mobility in Education? Evidence from Jordan. *World Bank Economic Review*, 32(3):633–655.

Azam, M. and Bhatt, V. (2015). Like Father, Like Son? Intergenerational Educational Mobility in India. *Demography*, 52(6):1929–1959.

Aziz, I., J. (1990). INPRES' Role in the Reduction of Interregional Disparity. *Asian Economic Journal*, 4(2):1–27.

Azomahou, T. T. and Yitbarek, E. (2021). Intergenerational Mobility in Education: Is Africa Different? *Contemporary Economic Policy*, Forthcoming.

Bau, N., Rotemberg, M., Shah, M., and Steinberg, B. (2020). Human capital investment in the presence of child labor. Technical report, National Bureau of Economic Research.

Bazzi, S., Hilmy, M., and Marx, B. (2020). Islam and the state: Religious education in the age of mass schooling. Technical report, National Bureau of Economic Research.

Becker, G. (1981). *A Treatise on the Family.* Harvard University Press.

Becker, G., Kominers, S. D., Murphy, K., and Spenkuch, J. (2015). A Theory of Intergenerational Mobility. MPRA Paper 66334, University Library of Munich, Germany.

Becker, G. S., Kominers, S. D., Murphy, K. M., and Spenkuch, J. L. (2018). A Theory of Intergenerational Mobility. *Journal of Political Economy*, 126(S1):7–25.

Behrman, J. R. and Deolalikar, A. B. (1995). Are There Differential Returns to Schooling by Gender? The Case of Indonesian Labour Markets. *Oxford Bulletin of Economics and Statistics*, 57(1):97–117.

Berman, Y. (2022). The long-run evolution of absolute intergenerational mobility. *American Economic Journal: Applied Economics*, 14(3):61–83.

Bjorklund, A. and Jantti, M. (2020). Intergenerational mobility, intergenerational effects, sibling correlations, and equality of opportunity: A comparison of four approaches. *Research in Social Stratification and Mobility*, 70:100455.

Black, S. E. and Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility.

In *Handbook of Labor Economics*, volume 4, chapter 16, pages 1487–1541. Elsevier.

Black, S. E., Devereux, P. J., Lundborg, P., and Majlesi, K. (2020). Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviours. *Review of Economic Studies*, 87(4):1683–1725.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *American economic review*, 95(1):437–449.

Blanden, J. and Machin, S. (2013). Educational Inequality and The Expansion of United Kingdom Higher Education. *Scottish Journal of Political Economy*, 60(5):597–598.

Breen, R. (2010). Educational expansion and social mobility in the 20th century. *Social Forces*, 89(2):365–388.

Calkins, R. and Sengupta (1992). Indonesia, Women in Development: A Strategy for Continued Progress. Report no. idp-112, east asia and pacific regional series, World Bank Group.

Card, D., Domnisoru, C., and Taylor, L. (2022). The intergenerational transmission of human capital: evidence from the golden age of upward mobility. *Journal of Labor Economics*, 40(S1):S39–S95.

Carneiro, P., García, I. L., Salvanes, K. G., and Tominey, E. (2021). Intergenerational Mobility and the Timing of Parental Income. *Journal of Political Economy*, 129(3):757–788.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Cholli, N. and Durlauf, S. (2022). Intergenerational Mobility. NBER Working Paper Series 29760, National Bureau of Economic Rsearch.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. L. (1966). *Equality of Educational Opportunity.* Government Printing Office, Washington, DC.

Currie, J. and Moretti, E. (2003). Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. *The Quarterly Journal of Economics*, 118(4):1495–1532.

Deolalikar, A. B. (1993). Gender Differences in the Returns to Schooling and in School

Enrollment Rates in Indonesia. *Journal of Human Resources*, 28(4):899–932.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91(4):795–813.

Duflo, E. (2004). The medium run effects of educational expansion: evidence from a large school construction program in Indonesia. *Journal of Development Economics*, 74(1):163–197.

Emran, M. S., Jiang, H., and Shilpi, F. (2021). Is Gender Destiny? Gender Bias and Intergenerational Educational Mobility in India. GLO Discussion Paper Series 807, Global Labor Organization (GLO).

Emran, M. S. and Shilpi, F. (2011). Intergenerational Occupational Mobility in Rural Economy: Evidence from Nepal and Vietnam. *Journal of Human Resources*, 46(2):427–458.

Emran, M. S. and Shilpi, F. (2015). Gender, Geography, and Generations: Intergenerational Educational Mobility in Post-Reform India. *World Development*, 72:362–380.

Emran, M. S. and Shilpi, F. (2021). Economic approach to intergenerational mobility: Measures, methods, and challenges in developing countries. In Iversen, V., Krishna, A., and Sen, K., editors, *Social mobility in developing countries: Concepts, methods, and determinants*. Oxford University Press.

Filmer, D. (2007). If you build it, will they come? School availability and school enrolment in 21 poor countries. *Journal of Development Studies*, 43(5):901–928.

Hanushek, E. A. (2002). Publicly provided education. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 4, chapter 30, pages 2045–2141. Elsevier.

Heckman, J. and Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics*, 6(1):689–733.

Heneveld, W. (1979). Indonesian education in the seventies: Problems of rapid growth. *Southeast Asian Affairs*, pages 142–154.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., and Verashchagina, A. (2008). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis & Policy*, 7(2):1–48.

Hilger, N. G. (2015). The great escape: Intergenerational mobility in the united states since

1940. Technical report, National Bureau of Economic Research.

Iversen, V., Krishna, A., and Sen, K. (2019). Beyond poverty escapes – social mobility in developing countries: A review article. *World Bank Research Observer*, 34(2):239–273.

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics*, 7(1):63–88.

Jung, D., Bharati, T., and Chin, S. (2021). Does Education Affect Time Preference? Evidence from Indonesia. *Economic Development and Cultural Change*, 69(4):1451–1499.

Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *Journal of Political Economy*.

Lockheed, M. E., Verspoor, A. M., et al. (1991). *Improving primary education in developing countries.* Oxford University Press for World Bank.

Lou, J. and Li, J. (2022). Export expansion and intergenerational education mobility: Evidence from china. *China Economic Review*, 73:101797.

Maasoumi, E., Wang, L., and Zhang, D. (2022). Generalized intergenerational mobility regressions. Working paper, Emory University.

Maccini, S. and Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review*, 99(3):1006–26.

Martinez-Bravo, M. (2017). The Local Political Economy Effects of School Construction in Indonesia. *American Economic Journal: Applied Economics*, 9(2):256–289.

Mazumder, B. (2005). Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. *The Review of Economics and Statistics*, 87(2):235–255.

Mazumder, B., Rosales-Rueda, M., and Triyana, M. (2019). Intergenerational Human Capital Spillovers: Indonesia's School Construction and Its Effects on the Next Generation. *AEA Papers and Proceedings*, 109:243–249.

Mogstad, M. and Torsvik, G. (2021). Family Background, Neighborhoods and Intergenerational Mobility. NBER Working Papers 28874, National Bureau of Economic Research, Inc.

Neidhofer, G., Serrano, J., and Gasparini, L. (2018). Educational inequality and intergenerational mobility in Latin America: A new database. *Journal of Development Economics*, 134(C):329–349.

Neilson, C. and Zimmerman, S. (2014). The effect of school construction on test scores, school enrollment, and home prices. *Journal of Public Economics*, 120:18–31.

Nicoletti, C. and Francesconi, M. (2006). Intergenerational mobility and sample selection in short panels. *Journal of Applied Econometrics*, 21(8):1265–1293.

Orazem, P. F. and King, E. M. (2008). Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy. In Schultz, T. P. and Strauss, J. A., editors, *Handbook of Development Economics*, volume 4, chapter 55, pages 3475–3559. Elsevier.

Parman, J. (2011). American Mobility and the Expansion of Public Education. *The Journal of Economic History*, 71(1):105–132.

Pekkarinen, T., Pekkala, S., and Uusitalo, R. (2009). School tracking and intergenerational income mobility: Evidence from the finnish comprehensive school reform. *Journal of Public Economics*, 93(7):965 – 973.

Pesaresia, M., Ehrlicha, D., Ferria, S., Florczyka, A., Freirea, S., M., F. H., Halkiaa, A., Juleaa, M., Kempera, T., and Soillea, P. (2015). Gloabl Human Settlement Snalysis for Disaster Risk Reduction. *36th International Symposium on Remote Sensing of Environment; International Society for Photogrammetry and Remote Sensing (ISPRS)*, pages 837–843.

Pfeffer, F. T. and Hertel, F. R. (2015). How has educational expansion shaped social mobility trends in the united states? *Social Forces*, 94(1):143–180.

Pitt, M. M., Rosenzweig, M. R., and Gibbons, D. M. (1993). The determinants and consequences of the placement of government programs in indonesia. *The World Bank Economic Review*, 7(3):319–348.

Salvanes, K. (2023). What are the Drivers Intergenerational Mobility? The Role of Family, Neighborhood, Education, and Social Class: A Review of Bukodi and Goldthorpe's Social Mobility and Education in Britain. *Journal of Economic Literature*. Forthcoming.

Scott, G. L. (1985). Indonesian Women and Development. Technical report.

Solon, G. (1992). Intergenerational Income Mobility in the United States. *American Economic Review*, 82(3):393–408.

Solon, G. (1999). Intergenerational mobility in the labor market. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, pages

1761–1800. Elsevier.

Tilak, J. (1993). East Asia. In Hiil, A. and King, E., editors, *Women's Education in Developing Countries: Barriers, Benefits, and Policies*, chapter 7, pages 247–279. World Bank.

Torche, F. (2013). How do we characteristically measure and analyze intergenerational mobility. *Stanford Center on Poverty and Inequality.*

Torche, F. (2015). Intergenerational Mobility and Equality of Opportunity. *European Journal of Sociology*, 56(3):343–371.

Torche, F. (2019). Educational mobility in developing countries. WIDER Working Paper Series 2019-88.

UIS (2022). Out-of-school rate (1 year before primary, primary education, lower secondary education, upper secondary education). Technical report, Montreal: UNESCO Institute for Statistics.

Vella, F. and Karmel, T. (1999). Evaluating the Impact of Educational Expansion on the Occupational Status of Youth. *Australian Economic Papers*, 38(3):310–327.

World Bank (2006). *World development report 2006: Equity and development.* The World Bank.

World Bank (2018). *World Development Report 2018: Learning to Realize Education's Promise.* Number 28340 in World Bank Publications - Books. The World Bank Group.

Yu, Y., Fan, Y., and Yi, J. (2020). The One-Child Policy Amplifies Economic Inequality across Generations in China. IZA Discussion Papers 13617, Institute of Labor Economics (IZA).

Table 1: **Summary statistics**

Panel A: Full Sample (Obs: 2,048,164)

	Mean (1)	SD (2)
Child's Edu	8.84	4.33
Child's Primary Completion(=1)	0.90	0.30
Father's Edu	5.28	4.52
Normalized Inpres Intensity	0.22	0.11
Enrollment Rate in 1971	0.17	0.08
Number of Children in 1971 (in thousands)	194.29	124.18
Allocation of Water and Sanitation 1973–1978	0.48	0.27

Panel B: Exposed and Comparison Cohort

	Mean (1)	SD (2)
Exposed Cohort (Born Between 1968-1972), Obs= 1,796,198		
Child's Edu	9.12	4.20
Child's Primary Completion(=1)	0.92	0.28
Father's Edu	5.43	4.54
Comparison Cohort (Born Between 1957 to 1962), Obs= 251,966		
Child's Edu	6.79	4.68
Child's Primary Completion(=1)	0.77	0.42
Father's Edu	4.20	4.26

Notes: The variable *Normalized Inpres Intensity* measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. The full sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Edu represents years of schooling, which was calculated based on the education level completed. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table 2: **Effects of Inpres Schools on Intergenerational Educational Mobility: Dependent Variable: Children's Years of Schooling**

	Linear Model		Quadratic Model	
	Sons	**Daughters (Relevant CEF)**	**Sons (Relevant CEF)**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	1.585***	0.513	1.534***	0.476
	(0.387)	(0.466)	(0.380)	(0.492)
Father's Edu × Born 1968-72 × Inpres	-0.129**	0.043	-0.255***	-0.020
	(0.054)	(0.062)	(0.086)	(0.099)
Father's Edu Sq × Born 1968-72 × Inpres			0.015**	0.008
			(0.007)	(0.008)
R2	0.322	0.409	0.323	0.409
Observations	1199814	848350	1199814	848350

Notes: The **Relevant CEFs** imply correct functional form, which are based on Table A.2 in the online appendix. Quadratic model is adopted following Becker et al. (2015). Robust standard errors are in parentheses, clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Sample corresponds to children born between 1957 and 1962 (comparison), or 1968 to 1972 (treatment). Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table. All coefficients are reported in Table A.1. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table 3: **Effects of Inpres Schools on Intergenerational Educational Mobility: Dependent Variable: Children's Primary Completion**

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons (Relevant CEF)**	**Daughters (Relevant CEF)**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.204***	0.216***	0.195***	0.197***
	(0.041)	(0.044)	(0.042)	(0.047)
Father's Edu × Born 1968-72 × Inpres	-0.022***	-0.014**	-0.038***	-0.017*
	(0.006)	(0.006)	(0.011)	(0.010)
Father's Edu Sq × Born 1968-72 × Inpres			0.002***	0.001*
			(0.001)	(0.001)
R2	0.107	0.172	0.117	0.188
Observations	1199814	848350	1199814	848350

Notes: The **Relevant CEFs** imply correct functional form, which are based on Table A.2 in the online appendix. Quadratic model is adopted following Becker et al. (2015). Robust standard errors are in parentheses, clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Sample corresponds to children born between 1957 and 1962 (comparison), or 1968 to 1972 (treatment). Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table. All coefficients are reported in Table A.3. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table 4: **Effects of Inpres Schools on Relative Mobility in Primary Completion of Children: Relative Mobility is Measured by Intergenerational Marginal Association (IGMA)**

	Sons				Daughters			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Highest Intensity(=1)		**Mean Intensity(=0.215)**		**Highest Intensity(=1)**		**Mean Intensity(=0.215)**	
		Normalized		**Normalized**		**Normalized**		**Normalized**
$\Delta IGMA_0$	-0.038	-157.95%	-0.008	-33.96%	-0.017	-70.66%	-0.004	-15.19%
	(0.011)		(0.002)		(0.010)		(0.002)	
$\Delta IGMA_6$	-0.013	-54.04%	-0.003	-11.62%	-0.004	-16.63%	-0.001	-3.57%
	(0.005)		-0.00108		(0.004)		(0.001)	
$\Delta IGMA_9$	-0.000	0.00%	0	0.00%	0.002	8.31%	0.000	1.79%
	(0.003)		(0.001)		(0.003)		(0.001)	
$\Delta IGMA_{12}$	0.013	54.04%	0.003	11.62%	0.009	37.41%	0.002	8.04%
	(0.004)		(0.001)		(0.005)		(0.001)	
$\Delta IGMA_{16}$	0.030	124.70%	0.006	26.81%	0.018	74.82%	0.004	16.09%
	(0.008)		(0.002)		(0.010)		(0.002)	

Notes: Robust standard errors are in parentheses, clustered at the district of birth(*** p<0.01, ** p<0.05, * p<0.1). IGMA is the slope of Conditional Expectation Function (CEF). $\Delta IGMA_y = \theta_7 + 2\theta_{11} E^p_{iy} \times Inpres$, where E^p_{iy} represents father's years of schooling for $y = 0, 6, 9, 12, 16$. The IGMA values are based on coefficients reported in Table 3. The variable Inpres intensity measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. The Normalized IGMA is the IGMA value relative to IGMA of children for father's with average years of education (5.57 years), comparison cohort (born between 1957 to 1962), and zero Inpres intensity for the full sample (combined sample of sons and daughters). Normalized IGMA values are reported in percentage. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table 5: **Effects of Inpres Schools on Absolute Mobility in Primary Completion of Children: Absolute Mobility is Measured by Expected Primary Completion (EPC)**

	Sons				Daughters			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Highest Intensity(=1)		**Mean Intensity(=0.215)**		**Highest Intensity(=1)**		**Mean Intensity(=0.215)**	
		Normalized		**Normalized**		**Normalized**		**Normalized**
ΔEPC_0	0.195	20.70%	0.042	4.45%	0.197	20.91%	0.042	4.50%
	(0.042)		(0.009)		(0.047)		(0.010)	
ΔEPC_6	0.042	4.46%	0.009	0.96%	0.132	14.01%	0.028	3.01%
	(0.033)		(0.007)		(0.034)		(0.007)	
ΔEPC_9	0.023	2.44%	0.005	0.52%	0.130	13.80%	0.028	2.97%
	(0.038)		(0.008)		(0.034)		(0.007)	
ΔEPC_{12}	0.042	4.46%	0.009	0.96%	0.147	15.61%	0.032	3.36%
	(0.040)		(0.009)		(0.032)		(0.007)	
ΔEPC_{16}	0.126	13.38%	0.027	2.88%	0.201	21.34%	0.043	4.59%
	(0.045)		(0.010)		(0.043)		(0.009)	

Notes: Robust standard errors are in parentheses, clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Absolute Mobility is measured by expected probability of primary completion (EPC) conditional on father's schooling. $\Delta EPC_y = \theta_4 \times Inpres + \theta_7 \times Inpres \times E^p_{iy} + \theta_{11} \times Inpres \times (E^p_{iy})^2$, where E^p_{iy} represents father's years of schooling for $y = 0, 6, 9, 12, 16$. EPC_y are calculated at y father's years of schooling, where $y = 0, 6, 9, 12, 16$. The EPC values are based on coefficients reported in Table 3. The variable Inpres intensity measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. The Normalized EPC is the EPC value relative to EPC of children for father's with average years of education (5.57 years), comparison cohort (born between 1957 to 1962), and zero Inpres intensity for the full sample (combined sample of sons and daughters). Normalized EPC values are reported in percentage. Data sources: Indonesia's full count census 2000 and Duflo (2001).

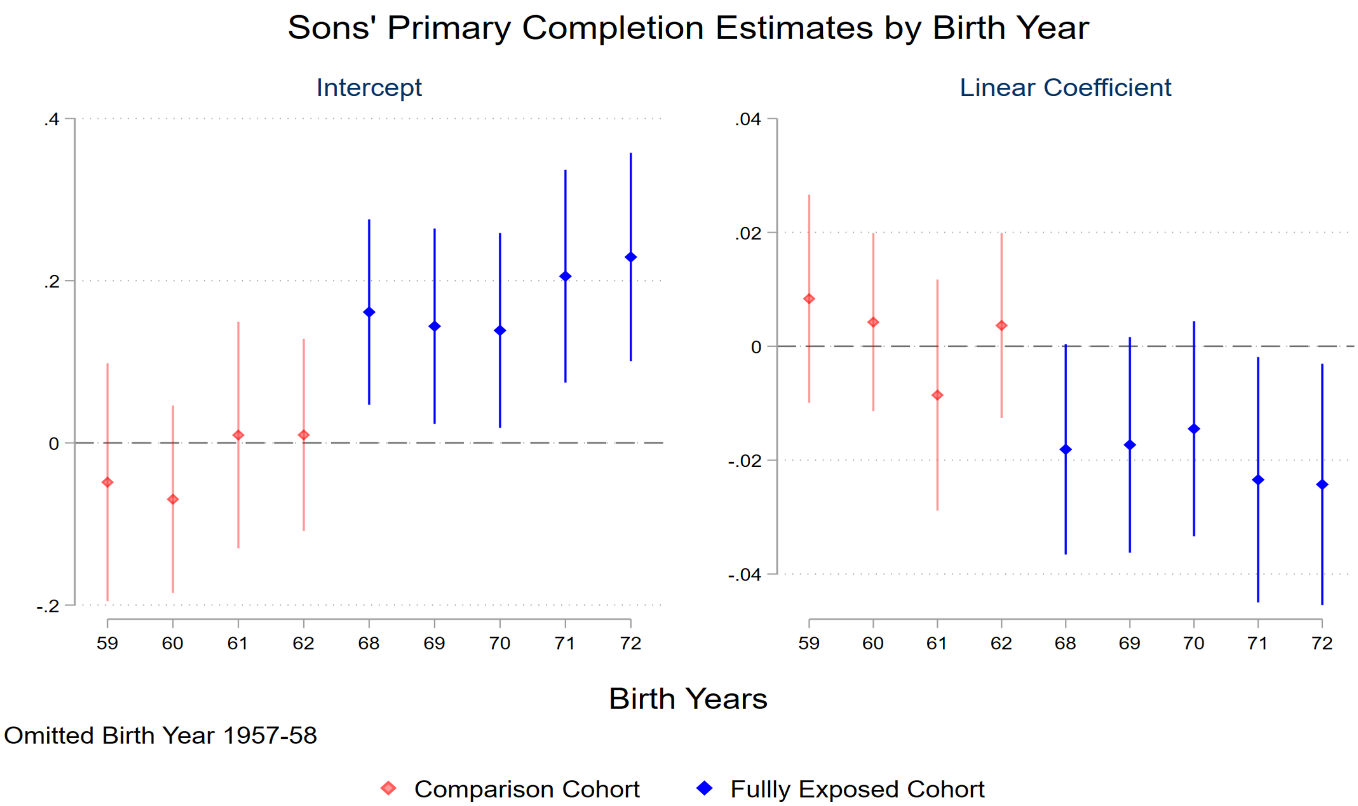
Table 6: **Evidence on Gender-based Crowding Out: Effects of Inpres on Senior High Completion**

	Panel A: Full Sample	
	Sons	**Daughters**
	(1)	(2)
Born 1968-72 × Inpres	0.076*	-0.084*
	(0.039)	(0.049)
Father's Edu × Born 1968-72 × Inpres	-0.006	0.008
	(0.011)	(0.011)
Father's Edu Sq × Born 1968-72 × Inpres	0.000	-0.000
	(0.001)	(0.001)
R2	0.270	0.343
Observations	1199814	848350

	Panel B: Matrilineal and Patrilineal Samples			
	Matrilineal		**Patrilineal**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	-0.749*	0.130	0.082**	-0.073
	(0.367)	(0.628)	(0.040)	(0.049)
Father's Edu × Born 1968-72 × Inpres	0.144	-0.071	-0.006	0.008
	(0.108)	(0.062)	(0.011)	(0.011)
Father's Edu Sq × Born 1968-72 × Inpres	-0.014**	0.003	0.000	-0.000
	(0.006)	(0.007)	(0.001)	(0.001)
R2	0.239	0.257	0.270	0.343
Observations	19901	18381	1179913	829969

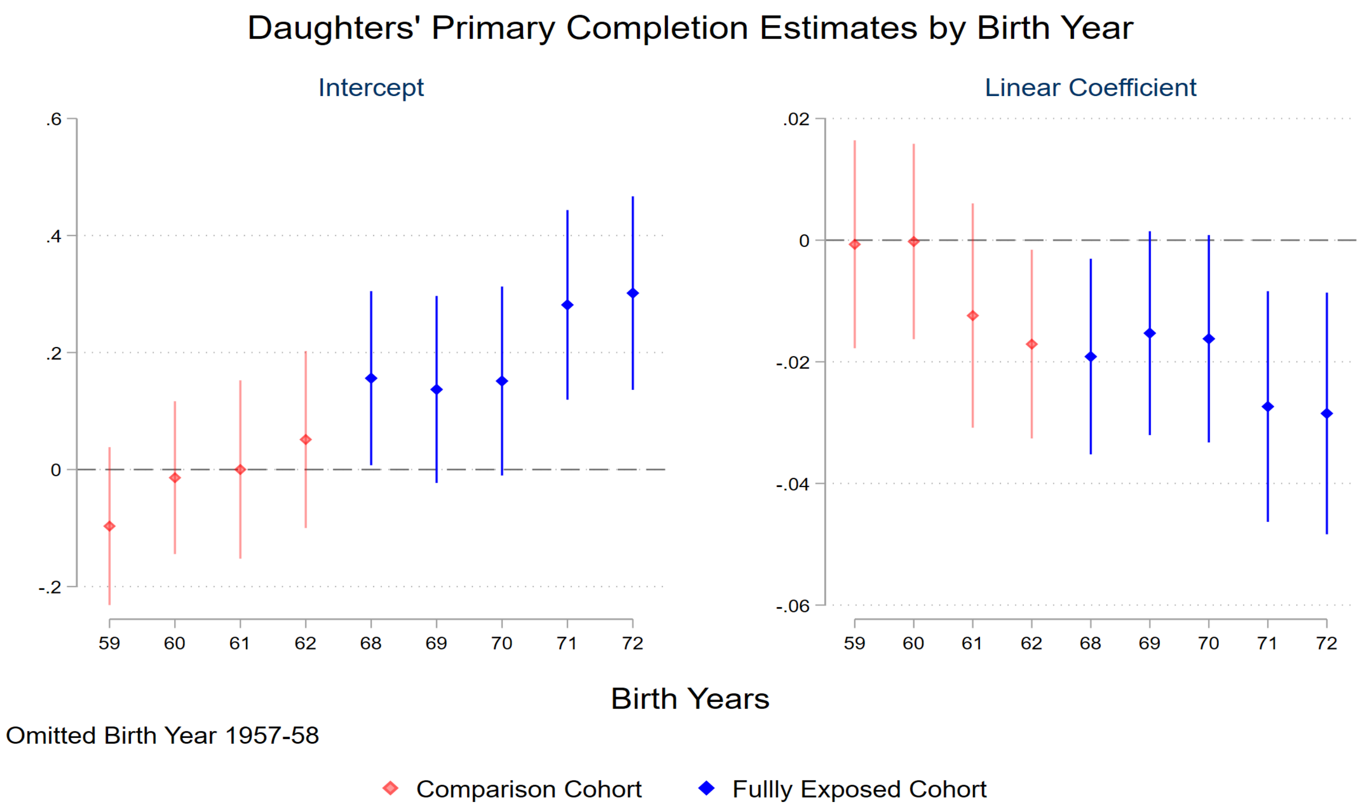
Notes: Robust standard errors are in parentheses, clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). We define West Sumatra as a matrilineal society and the rest of Indonesia as a Patrilineal society. Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Covariates include birth FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Senior high completion takes the value of 1 if the child has completed senior high and 0, otherwise. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table; full table, with all coefficients, is available upon request. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Figure 1: Event Study of Inpres Impacts with Linear CEF for Sons



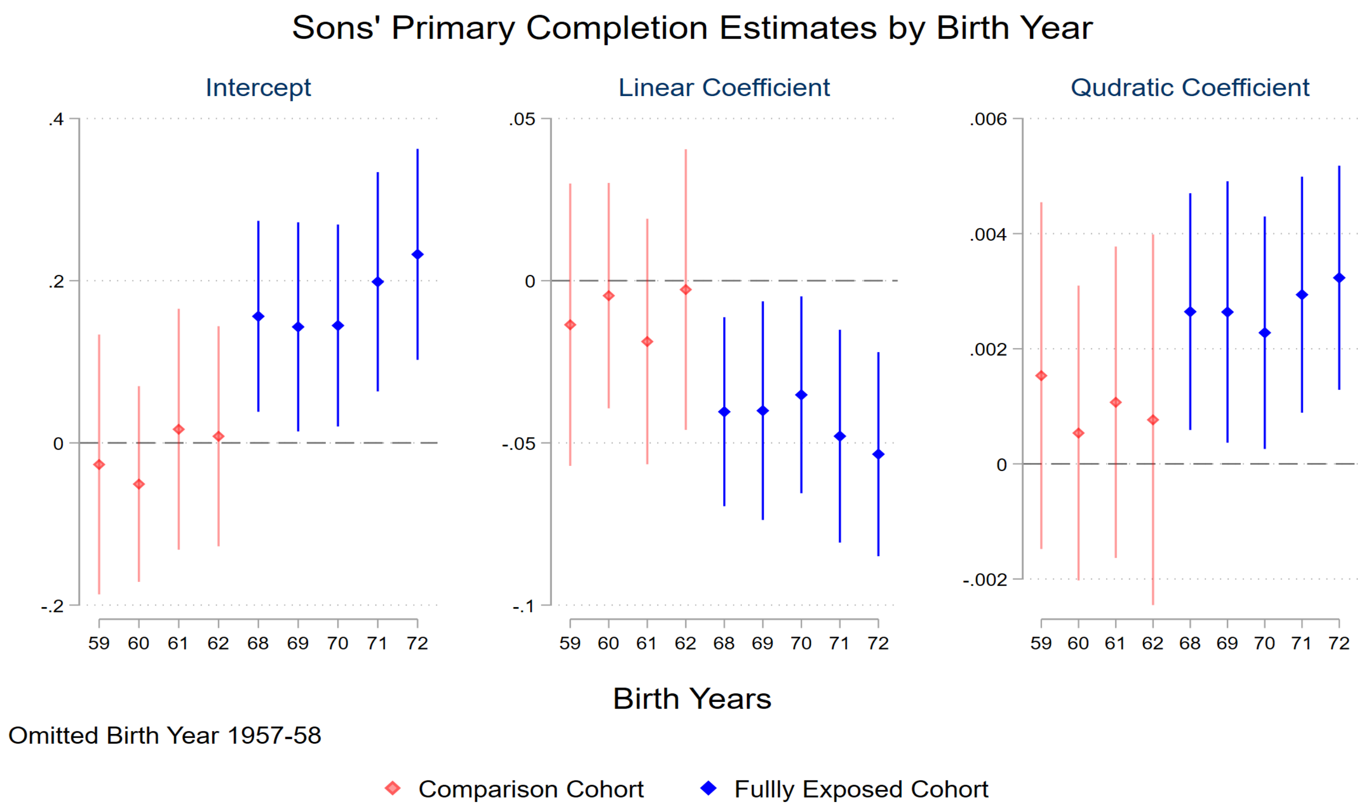
Notes: This figure plots the estimates of β_4 and β_7 of equation (2) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The estimates of partially exposed cohorts along with comparison cohorts and fully exposed cohorts are plotted in **Figure A1**. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001).

Figure 2: Event Study of Inpres Impacts with Linear CEF for Daughters



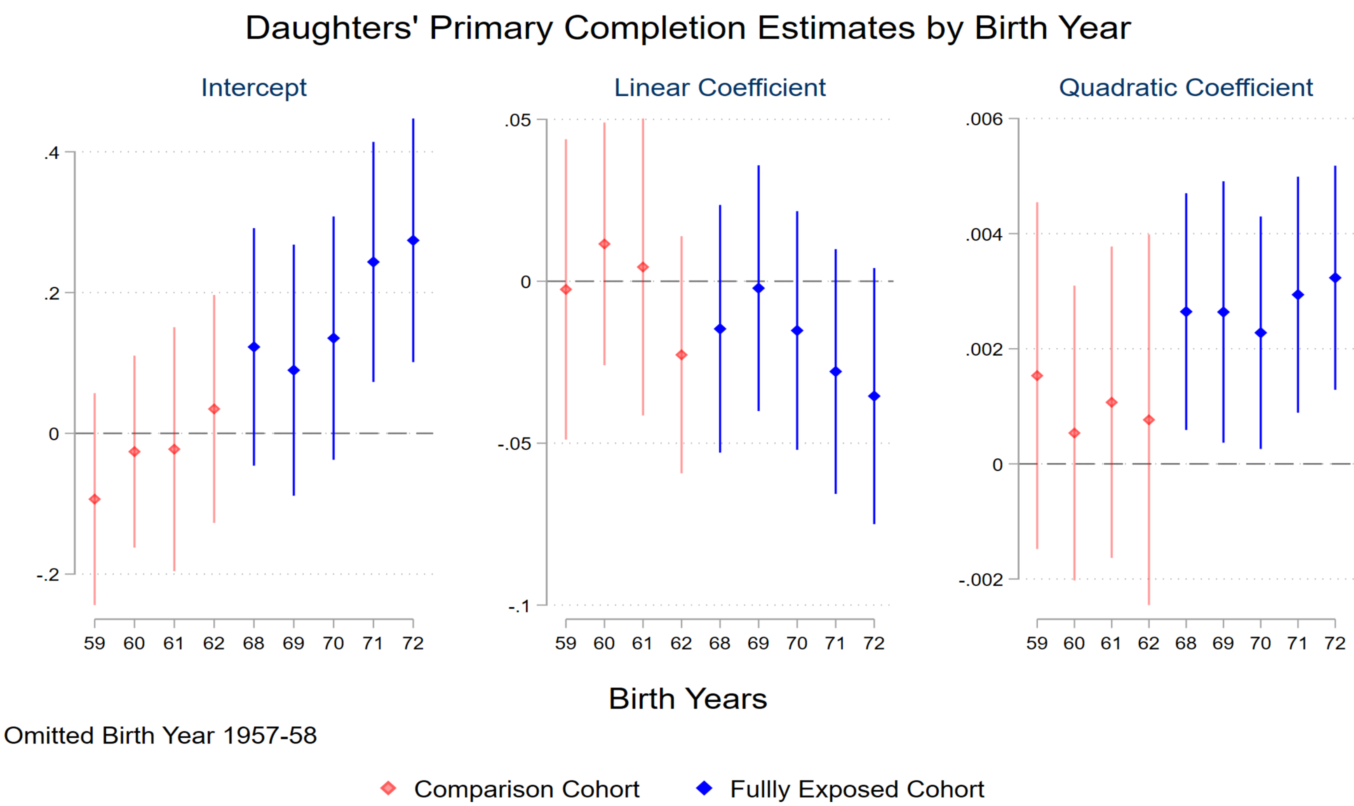
Notes: This figure plots the estimates of β_4 and β_7 of equation (2) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The estimates of partially exposed cohorts along with comparison cohorts and fully exposed cohorts are plotted in **Figure A2**. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001).

Figure 3: Event Study of Inpres Impacts with Quadratic CEF for Sons



Notes: This figure plots the estimates of θ_4, θ_7, and θ_{11} of equation (3) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The estimates of partially exposed cohorts along with comparison cohorts and fully exposed cohorts are plotted in **Figure A3**. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001).

Figure 4: Event Study of Inpres Impacts with Quadratic CEF for Daughters



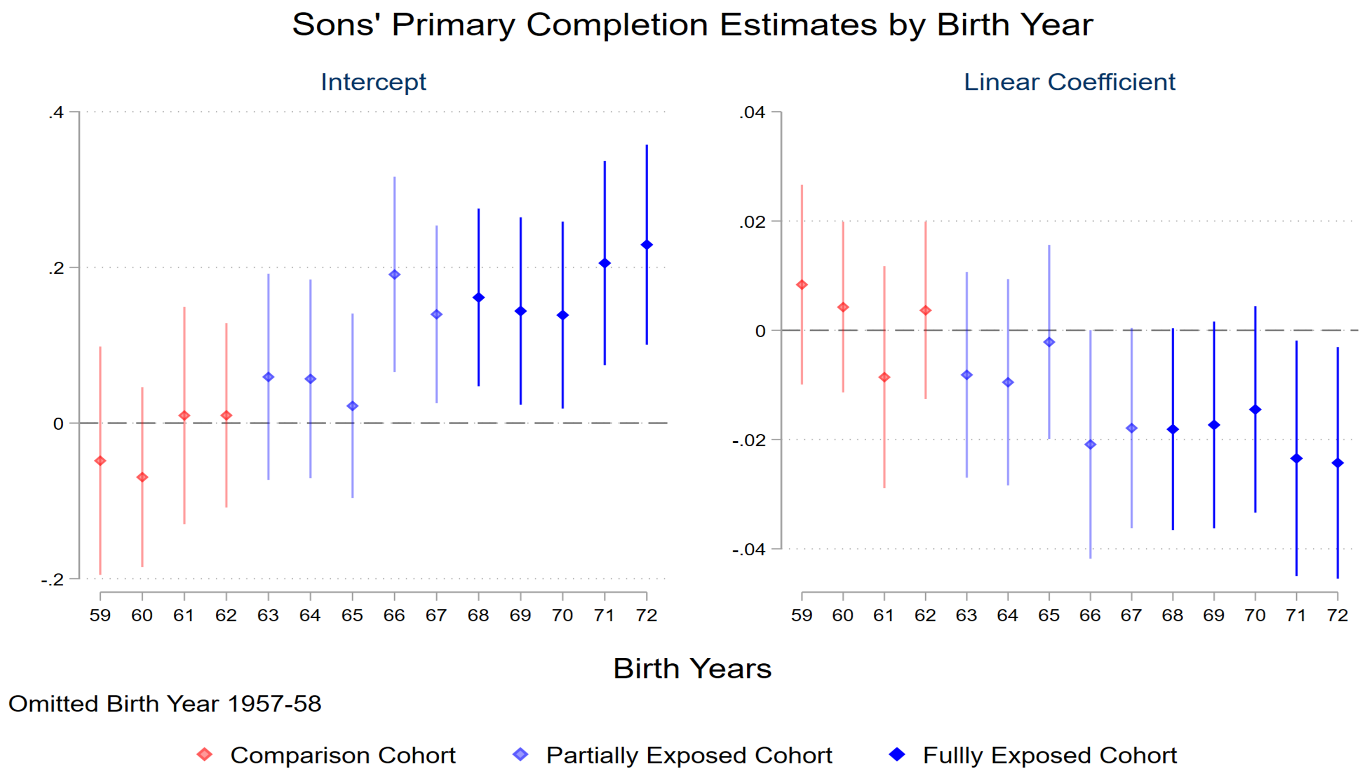
Notes: This figure plots the estimates of θ_4, θ_7, and θ_{11} of equation (3) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The estimates of partially exposed cohorts along with comparison cohorts and fully exposed cohorts are plotted in **Figure A4**. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001).

NOT FOR PUBLICATION

Online Appendix to:
"Public Primary School Expansion, Gender-Based Crowding Out, and Intergenerational Educational Mobility"

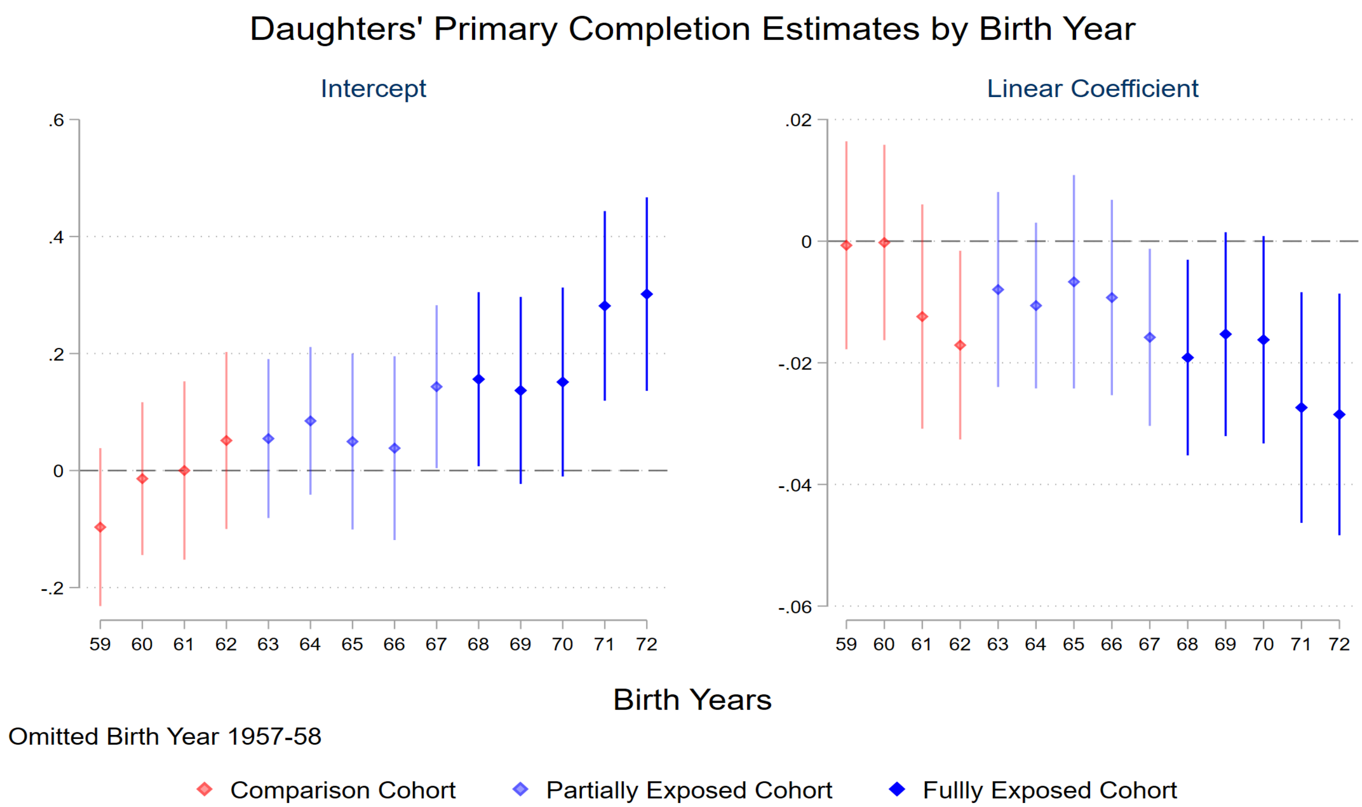
Online Appendix A

Figure A1: Event Study of Inpres Impacts with Linear CEF for Sons
Including Partially Exposed Cohorts



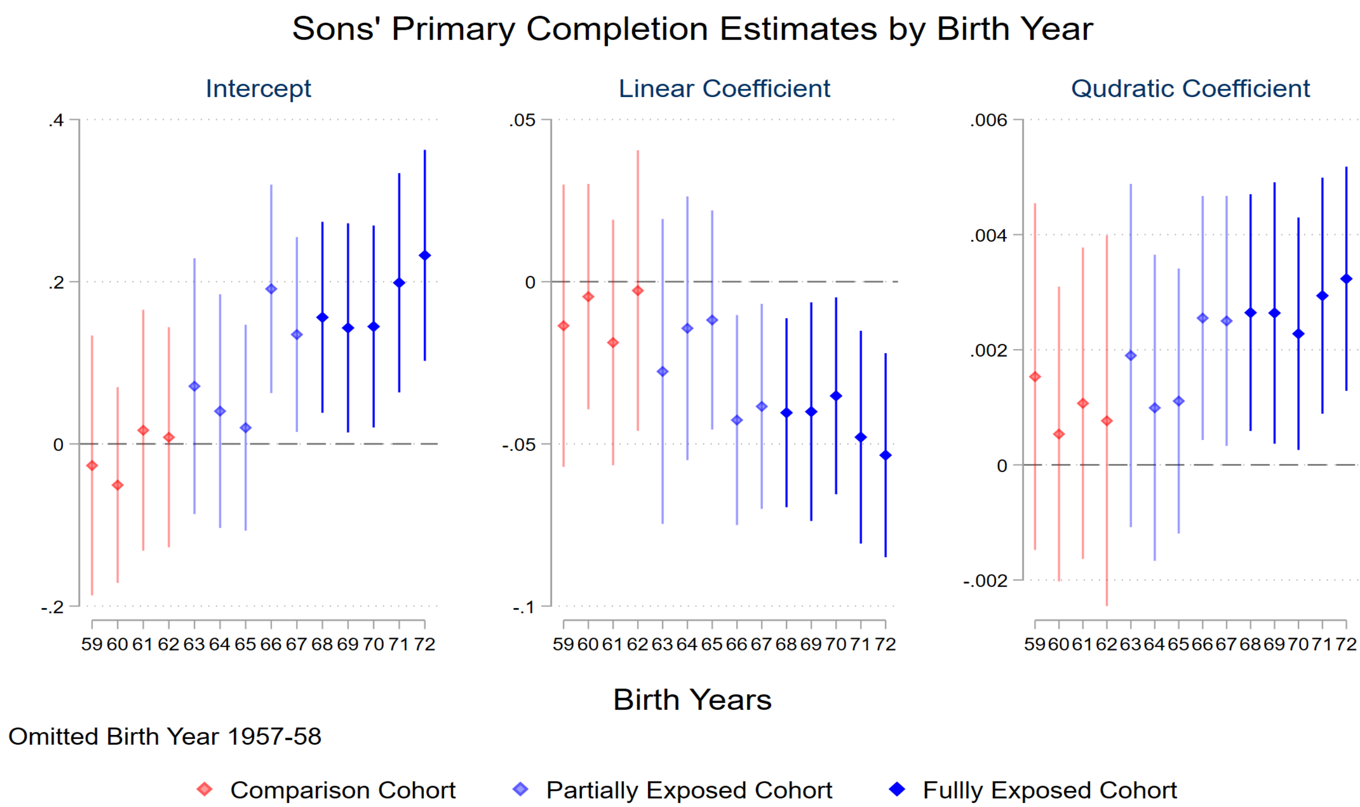
Notes: This figure plots the estimates of β_4 and β_7 of equation (2) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001).

Figure A2: Event Study of Inpres Impacts with Linear CEF for Daughters Including Partially Exposed Cohorts



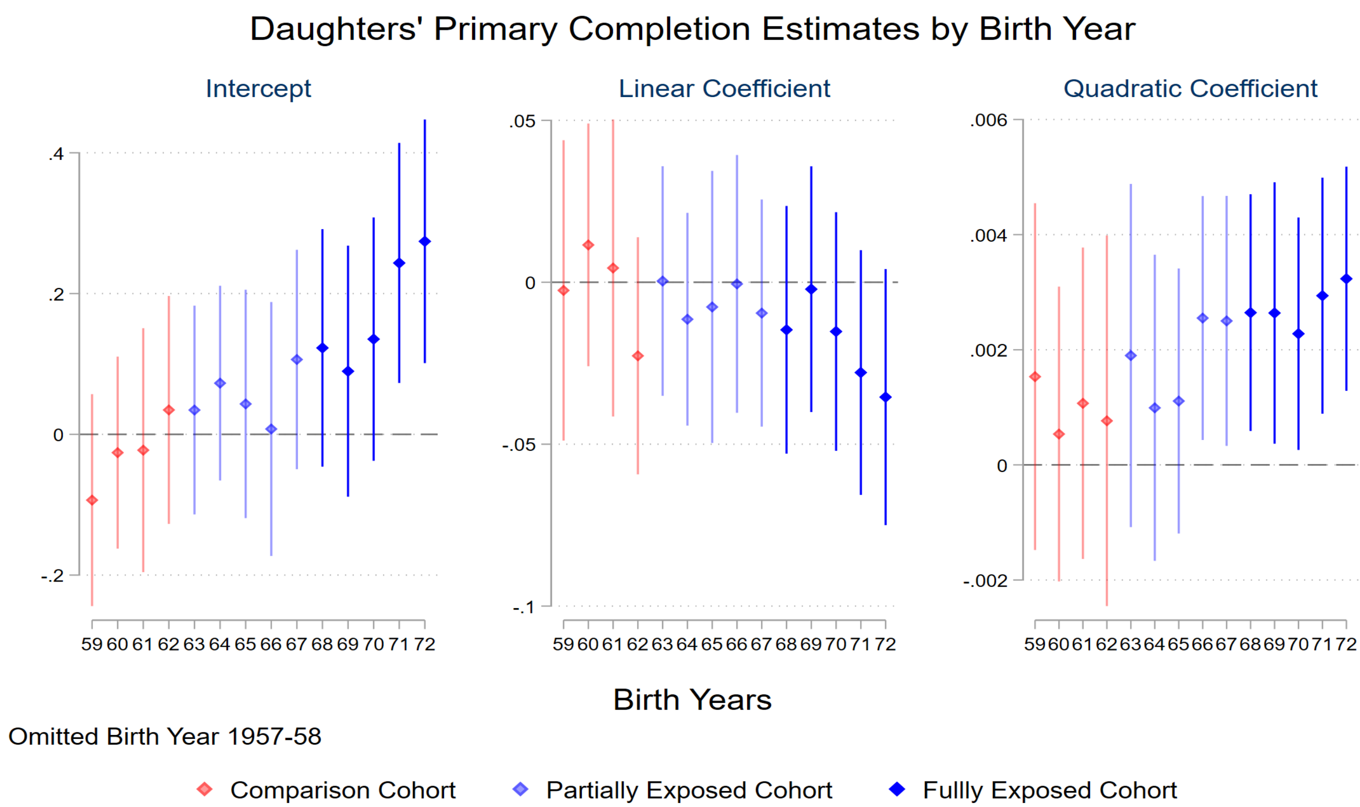
Notes: This figure plots the estimates of β_4 and β_7 of equation (2) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001).

Figure A3: Event Study of Inpres Impacts with Quadratic CEF for Sons Including Partially Exposed Cohorts



Notes: This figure plots the estimates of θ_4, θ_7, and θ_{11} of equation (3) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth$\times$1971 enrollment, year of birth$\times$1971 number of children, year of birth$\times$water sanitation program, year of birth dummies, following Duflo (2001).

Figure A4: Event Study of Inpres Impacts with Quadratic CEF for Daughters Including Partially Exposed Cohorts



Notes: This figure plots the estimates of θ_4, θ_7, and θ_{11} of equation (3) in the manuscript by birth year. For each birth year, the diamond symbol represents the value of the estimate with the confidence interval at 95 percent level. Sample corresponds to individuals born between 1957 to 1972. The Inpres program started in 1973-74. Therefore, children born between 1968 to 1972 are considered fully exposed, children born between 1963 to 1967 are considered partially exposed, and children born between 1957 to 1963 are comparison cohorts. The omitted birth year cohorts are children born in 1957 and 1958. The dependent variable takes the value of 1 if the child has completed primary or more and 0 otherwise. Robust standard errors are in parentheses, clustered at the district of birth. Covariates include birth district FE, year of birth$\times$1971 enrollment, year of birth$\times$1971 number of children, year of birth$\times$water sanitation program, year of birth dummies, following Duflo (2001)

Table A.1: **Effects of Inpres Schools on Intergenerational Educational Mobility: Dependent Variable: Children's Years of Schooling (Full Table)**

Dependent Variable: Years of Schooling				
	Linear Model		**Quadratic Model**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Father's Edu	0.450***	0.528***	0.368***	0.482***
	(0.018)	(0.018)	(0.033)	(0.035)
Born 1968-72 × Inpres	1.585***	0.513	1.534***	0.476
	(0.387)	(0.466)	(0.380)	(0.492)
Father's Edu × Born 1968-72	-0.053***	-0.083***	-0.081***	-0.089***
	(0.013)	(0.017)	(0.021)	(0.024)
Father's Edu × Inpres	0.257***	0.113*	0.561***	0.311**
	(0.076)	(0.063)	(0.138)	(0.132)
Father's Edu × Born 1968-72 × Inpres	-0.129**	0.043	-0.255***	-0.020
	(0.054)	(0.062)	(0.086)	(0.099)
Father's Edu Sq			0.007***	0.004
			(0.002)	(0.003)
Father's Edu Sq × Born 1968-72			0.001	-0.000
			(0.002)	(0.002)
Father's Edu Sq × Inpres			-0.027***	-0.018*
			(0.008)	(0.010)
Father's Edu Sq × Born 1968-72 × Inpres			0.015**	0.008
			(0.007)	(0.008)
R2	0.322	0.409	0.323	0.409
Observations	1199814	848350	1199814	848350

Notes: Robust standard errors are in parentheses, clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table A.2: **Quadratic CEFs of Exposed and Comparison Cohorts**

Panel A: Dependent Variable: Years of Schooling of Children

	Sons		**Daughters**	
	Comparison	**Exposed**	**Comparison**	**Exposed**
	(1)	(2)	(3)	(4)
Father's Edu	0.532***	0.388***	0.627***	0.530***
	(0.02294)	(0.01835)	(0.02688)	(0.02390)
Father's Edu Sq	0.002	0.007***	0.000	0.001
	(0.00139)	(0.00110)	(0.00173)	(0.00140)
Constant	5.044***	6.837***	3.651***	5.957***
	(0.09773)	(0.08767)	(0.09058)	(0.09836)
R2	0.260	0.267	0.334	0.330
Observations	119825	1079989	132141	716209

Panel B: Dependent Variable: Primary Completion of Children

	Sons		**Daughters**	
	Comparison	**Exposed**	**Comparison**	**Exposed**
	(1)	(2)	(3)	(4)
Father's Edu	0.059***	0.028***	0.082***	0.040***
	(0.00220)	(0.00136)	(0.00262)	(0.00197)
Father's Edu Sq	-0.003***	-0.001***	-0.004***	-0.002***
	(0.00012)	(0.00006)	(0.00013)	(0.00009)
Constant	0.663***	0.837***	0.530***	0.771***
	(0.01075)	(0.00720)	(0.01282)	(0.01004)
R2	0.120	0.059	0.188	0.094
Observations	119825	1079989	132141	716209

Notes: Robust standard errors are in parentheses, clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Quadratic model is adopted following Becker et al. (2015). Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. Comparison cohorts are children born between 1957 and 1962, and exposed cohorts are children born between 1968 to 1972. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Data source: Indonesia's full count census 2000.

Table A.3: **Effects of Inpres Schools on Intergenerational Educational Mobility: Dependent Variable: Children's Primary School Completion (Full Table)**

	Dependent Variable: Primary Completion			
	Linear Model		**Quadratic Model**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Father's Edu	0.017***	0.028***	0.037***	0.060***
	(0.003)	(0.003)	(0.004)	(0.005)
Born 1968-72 × Inpres	0.204***	0.216***	0.195***	0.197***
	(0.041)	(0.044)	(0.042)	(0.047)
Father's Edu × Born 1968-72	-0.010***	-0.019***	-0.022***	-0.037***
	(0.001)	(0.002)	(0.003)	(0.002)
Father's Edu × Inpres	0.043***	0.044***	0.083***	0.076***
	(0.010)	(0.013)	(0.018)	(0.021)
Father's Edu × Born 1968-72 × Inpres	-0.022***	-0.014**	-0.038***	-0.017*
	(0.006)	(0.006)	(0.011)	(0.010)
Father's Edu Sq			-0.002***	-0.003***
			(0.000)	(0.000)
Father's Edu Sq × Born 1968-72			0.001***	0.002***
			(0.000)	(0.000)
Father's Edu Sq × Inpres			-0.004***	-0.004***
			(0.001)	(0.001)
Father's Edu Sq × Born 1968-72 × Inpres			0.002***	0.001*
			(0.001)	(0.001)
R2	0.107	0.172	0.117	0.188
Observations	1199814	848350	1199814	848350

Notes: Robust standard errors are in parentheses, clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table A.4: **Effects of Inpres Schools on Relative and Absolute Mobility in Years of Schooling**

Panel A: Intergenerational Mobility Association (IGMA)/ Relative Mobility (Sons)

	(1)	(2)	(3)	(4)
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
		Normalized		**Normalized**
$\Delta IGMA_0$	-0.255	-51.77%	-0.055	-11.13%
	(0.086)		(0.018)	
$\Delta IGMA_6$	-0.079	-16.04%	-0.017	-3.45%
	(0.052)		(0.011)	
$\Delta IGMA_9$	0.009	1.83%	0.002	0.39%
	(0.070)		(0.015)	
$\Delta IGMA_{12}$	0.097	19.69%	0.021	4.23%
	(0.101)		(0.022)	
$\Delta IGMA_{16}$	0.215	43.65%	0.046	9.38%
	(0.149)		(0.032)	

Panel B: Expected Schooling (ES)/ Absolute Mobility (Sons)

	(1)	(2)	(3)	(4)
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
		Normalized		**Normalized**
ΔES_0	1.534	17.52%	0.330	3.77%
	(0.380)		(0.082)	
ΔES_6	0.533	6.09%	0.115	1.31%
	(0.347)		(0.075)	
ΔES_9	0.428	4.89%	0.092	1.05%
	(0.390)		(0.084)	
ΔES_{12}	0.589	6.73%	0.127	1.45%
	(0.515)		(0.111)	
ΔES_{16}	1.214	13.86%	0.261	2.98%
	(0.895)		(0.192)	

Notes: Robust standard errors are in parentheses, clustered at the district of birth(*** p<0.01, ** p<0.05, * p<0.1). Intergenerational Mobility Association(IGMA) is the slope of Conditional Expectation Function (CEF). Here, $\Delta IGMA_y = \theta_7 + 2\theta_{11} E^p_{iy} \times Inpres$, where E^p_{iy} represents father's years of schooling for $y = 0, 6, 9, 12, 16$. Absolute Mobility is measured by expected schooling conditional on father's schooling. Here, $\Delta ES_y = \theta_4 \times Inpres + \theta_7 \times Inpres \times E^p_{iy} + \theta_{11} \times Inpres \times (E^p_{iy})^2$, where E^p_{iy} represents father's years of schooling for $y = 0, 6, 9, 12, 16$. The IGMA and ES values are based on coefficients reported in Table 2. The variable Inpres intensity measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. The Normalized IGMA is the IGMA value relative to IGMA of children for father's with average years of education (5.57 years), comparison cohort (born between 1957 to 1962), and zero Inpres intensity for the full sample (combined sample of sons and daughters). Normalized IGMA values are reported in percentage. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table A.5: **Quadratic CEFs of Exposed and Comparison Cohorts for Senior High School Completion**

Dependent Variable: Senior High Completion of Children				
	Sons		**Daughters**	
	Comparison	**Exposed**	**Comparison**	**Exposed**
	(1)	(2)	(3)	(4)
Father's Edu	0.0254***	0.0372***	0.0165***	0.0467***
	(0.00265)	(0.00278)	(0.00292)	(0.00326)
Father's Edu Sq	0.0022***	0.0013***	0.0028***	0.0008***
	(0.00019)	(0.00020)	(0.00021)	(0.00024)
Constant	0.1155***	0.1935***	0.0467***	0.1412***
	(0.00660)	(0.00790)	(0.00331)	(0.00709)
R2	0.213	0.226	0.261	0.284
Observations	119825	1079989	132141	716209

Notes: Robust standard errors in parentheses clustered at the district of birth (*** $p<0.01$, ** $p<0.05$, * $p<0.1$). Quadratic model is adopted following Becker et al. (2015). Senior high completion takes the value of 1 if a respondent has completed senior high or more schooling and 0 otherwise. Comparison cohorts are children born between 1957 and 1962, and exposed cohorts are children born between 1968 to 1972. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. Data source: Indonesia's full count census 2000.

Table A.6: **Alternative Models of Intergenerational Educational Mobility**

Panel A: Normalized Years of Schooling

	Linear Model		**Quadratic Model**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.382***	0.113	0.370***	0.104
	(0.093)	(0.102)	(0.091)	(0.108)
(Father's Edu /Sub-Group Sd) × Born 1968-72 × Inpres	-0.139**	0.044	-0.274***	-0.020
	(0.058)	(0.063)	(0.093)	(0.100)
(Father's Edu /Sub-Group Sd) Sq × Born 1968-72 × Inpres			0.071**	0.037
			(0.031)	(0.038)
R2	0.322	0.409	0.323	0.409
Observations	1199814	848350	1199814	848350

Panel B: Rank in Schooling Distribution

	Linear Model		**Quadratic Model**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.087***	-0.032	0.085**	-0.024
	(0.031)	(0.042)	(0.033)	(0.047)
Father's Edu (Mid) Nat. Rank × Born 1968-72 × Inpres	-0.084	0.104	-0.132	0.015
	(0.051)	(0.080)	(0.152)	(0.168)
Father's Edu (Mid) Nat. Rank Sq × Born 1968-72 × Inpres			0.075	0.099
			(0.157)	(0.175)
R2	0.333	0.424	0.341	0.430
Observations	1199814	848350	1199814	848350

Notes: Robust standard errors in parentheses clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Quadratic model is adopted following Becker et al. (2015). Family background is measured by father's schooling. Father's Edu is father's years of schooling. The normalized schooling measure was constructed for sons and daughters separately. The Sub-Group SD refers to standard deviation of father's years of schooling for that specific sub-samples (sons/daughters). The rank in schooling is constructed based on combined sample (Sons+Daughters), following Chetty et al. (2014). Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Covariates include birth district district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table; full table, with all coefficients, is available upon request. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Online Appendix B

(OB.1) Robustness Checks

(OB.1.1) Mother's Education

A central finding from our analysis is that there are no significant effects of Inpres schools on final educational attainment of girls (measured by years of schooling) even though the effects at the primary level are substantial. One might wonder whether the conclusions would be different if we used mother's education in place of father's education as a measure of family background of children. There is substantial evidence that the intergenerational link is much stronger between mothers and daughters (see, for example, Smith and Smith (2013), Emran and Shilpi (2011)). We report the estimates of equations (2) and (3) using mother's education in Table B.1 below. The evidence is very similar to what we found earlier in Table 2 and Table 3 using father's education. Perhaps, most important, there are no effects of Inpres schools on educational mobility of daughters when completed years of schooling is the measure of educational attainment. But, consistent with the evidence in Table 2 and Table 3, there are substantial and statistically significant effects at the primary level.

(OB.1.2) Alternative Comparison Groups, and Partially Exposed Groups

The DiD design in Table 2 and Table 3 uses data on 6 years from the pre-Inpres period to define the comparison groups. One might worry that the children from the earlier birth cohorts, say born in 1957-1958, are likely to be less comparable to the treatment birth cohorts (1968-1972) exposed to Inpres schools in the 1970s. Since we have a large data set, we can estimate the DiD model using only the more recent birth cohorts from the pre-Inpres sample without worrying about the loss of power. We estimated equations (2) and (3) using a number of such cutoffs to define the pre-Inpres comparison sample. We report the estimates for these alternative samples in Tables B.2-B.4 below in this appendix. The results are robust across these alternative pre-Inpres cohorts. We also provide estimated effects for the partially exposed cohorts (1963-1967): please see Table B.5. The estimated effects are numerically much smaller as one would expect, and many are not significant at the 10 percent level.

(OB.2) Evidence on Potential Coresidency Bias

(OB.2.1) Are the Census Estimates of Mobility CEFs Substantially Biased?

Recent evidence shows that sample truncation due to nonrandomly missing children biases the estimated slope parameter (called IGRC in the literature) downward in a *linear* educational mobility model, see Emran et al. (2018) and Azam and Bhatt (2015). To understand whether the estimates from the census data are substantially biased, we take advantage of the fact that the IFLS collected information on the nonresident children. If the estimates from census data are substantially biased downward, the estimates of relative mobility would be substantially different from those based on IFLS. Table B.6 below presents the estimates of the linear (see columns 1 and 2) and quadratic (see columns 4 and 5) mobility models for the census and IFLS data (including IFLS-East). The evidence suggests that the census estimates do not suffer from any substantial biases. For example, for years of schooling, the IGRC estimates are 0.489 (census) and 0.522 (IFLS) for girls, and 0.433 (census) and 0.455 (IFLS) for boys. Thus the estimated intergenerational persistence in the census data is smaller in magnitude which is consistent with the findings of Emran et al. (2018), but the magnitude of the downward bias is small. The estimates for primary completion similarly suggest small to moderate bias.[1] The evidence on the quadratic mobility model (for years of schooling and primary completion) also suggests that, in general, the census estimates are moderately biased downward, except for the case of sons' years of schooling for which the biases are somewhat larger.

(OB.2.2) Inverse Probability Weighting for Correction of Coresidency Bias

Given the evidence that the estimates of the mobility models using census are in general close to the IFLS estimates, it seems unlikely that our conclusions about the causal effects of Inpres schools can be driven by coresidency bias. But one can argue that we do not know the direction of bias for the causal effects from a priori considerations, and additional evidence would be helpful. To address any remaining concerns, we use inverse probability weighting to correct for possible biases in our estimates of the causal effects of Inpres. Nicoletti and Francesconi (2006) provides evidence in favor of IPW for correcting the coresidency bias in intergenerational mobility analysis relative to Heckman selection correction.

[1] For son's primary completion, the estimated slope parameter is identical across census and IFLS data.

Since house rental cost is likely to be the largest expense when children decide to leave parental home, we would expect coresidency rates to be higher in locations where rental rates are higher (Nicoletti and Francesconi (2006)). We take advantage of the recently available built up density data as a source of variation in house rent. Note that built-up density is a measure of supply of housing in a location, with a lower house rent where built-up density (supply) is higher. This implies that we expect a negative relation between built up density and coresideny rates on a priori grounds. The source of built-up data is the Global Human Settlement Layer (GHSL) (Pesaresia et al. (2015)). The built-up data are at 300 meters by 300 meters grids. We super-impose the digital maps from the censuses on the pixel-level data to estimate the total built-up area at the district level.

We use 1975 built up density in a district interacted with time trend in the estimating equation where the dependent variable is a dummy indicating whether a child is coresident with his/her father (with location and year fixed effects along with the other control variables used in the DiD design), and generate the estimated probability weights.[2] Built up density interacted with the time trend has substantial explanatory power in the coresidency equation; the estimated coefficient is significant at the 5 percent level, and has a negative sign consistent with the a priori expectations. To check if the IPW correction is effective, we estimate the linear and quadratic mobility models in Table B.6 discussed above using IPW correction for census data. The IPW corrected census estimates are reported in columns 3 (linear CEF) and column 6 (quadratic CEF) of Table B.6 in this appendix. The evidence suggests that the IPW correction makes the census estimates closer to the corresponding IFLS estimates.

We then apply the IPW correction to our estimates of the effects of Inpres schools reported in Table 2 and Table 3. The IPW corrected estimates for the relevant CEFs are reported in the appendix Table B.7.[3] A comparison with our main estimates in Table 2 and Table 3 shows that the inverse probability weighted estimates are not substantially different from the unweighted estimates; the null hypothesis of equality cannot be rejected for any of the causal estimates. The main conclusions of the paper thus remain intact after correction of possible

[2]The location fixed effects mop up any direct effect of built up density on children's education through agglomeration channel. The year fixed effects account for macroeconomic and international shocks common to all districts.

[3]Recall that the relevant CEF is linear for years of schooling in the case of daughters, but concave in all other cases.

biases due to sample truncation.

We check a plausible explanation for the evidence that IPW corrected estimates are not substantially different. If coresidency rates across districts are not significantly correlated with Inpres treatment intensity, then we expect IPW corrected estimates of the causal effects to be similar to the unweighted estimates in Table 2 and Table 3. For this exercise, we estimate the effects of Inpres schools on coresidency rates across districts using the same DiD design adopted in our main empirical analysis. The estimates are reported in Table B.8 below. The coefficient of Inpres treatment intensity is not significant at the 10 percent level for both sons and daughters. This suggests that sample truncation from coresidency in the census data do not bias the estimated causal effects in any significant manner. This is consistent with the evidence of Akresh et al. (2018) who find that correction of coresidency bias does not have any substantial effect on their estimates of intergenerational effects of Inpres schools (effects on the children of mothers exposed to Inpres while school aged).

References

Akresh, R., Halim, D., and Kleemans, M. (2018). Long-term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia. NBER Working Papers 25265, National Bureau of Economic Research, Inc.

Azam, M. and Bhatt, V. (2015). Like Father, Like Son? Intergenerational Educational Mobility in India. *Demography*, 52(6):1929–1959.

Emran, M. S., Greene, W., and Shilpi, F. (2018). When Measure Matters: Coresidency, Truncation Bias, and Intergenerational Mobility in Developing Countries. *Journal of Human Resources*, 53(3):589–607.

Emran, M. S. and Shilpi, F. (2011). Intergenerational Occupational Mobility in Rural Economy: Evidence from Nepal and Vietnam. *Journal of Human Resources*, 46(2):427–458.

Nicoletti, C. and Francesconi, M. (2006). Intergenerational mobility and sample selection in short panels. *Journal of Applied Econometrics*, 21(8):1265–1293.

Pesaresia, M., Ehrlicha, D., Ferria, S., Florczyka, A., Freirea, S., M., F. H., Halkiaa, A., Juleaa, M., Kempera, T., and Soillea, P. (2015). Gloabl Human Settlement Snalysis for

Disaster Risk Reduction. *36th International Symposium on Remote Sensing of Environment; International Society for Photogrammetry and Remote Sensing (ISPRS)*, pages 837–843.

Smith, G. and Smith, M. H. (2013). Like mother, like daughter? an economic comparison of immigrant mothers and their daughters. *International Migration*, 51(2):181–190.

Table B.1: **Effects of Inpres Schools on Intergenerational Educational Mobility: Mother's Years of Schooling as Family Background**

Panel A: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.184***	0.226***	0.180***	0.212***
	(0.044)	(0.043)	(0.044)	(0.044)
Mother's Edu × Born 1968-72 × Inpres	-0.021**	-0.020***	-0.041***	-0.023*
	(0.008)	(0.007)	(0.014)	(0.012)
Mother's Edu Sq × Born 1968-72 × Inpres			0.003***	0.001
			(0.001)	(0.001)
R2	0.099	0.157	0.107	0.169
Observations	1083281	738202	1083281	738202

Panel B: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	1.603***	0.755	1.592***	0.720
	(0.401)	(0.462)	(0.396)	(0.487)
Mother's Edu × Born 1968-72 × Inpres	-0.129*	0.010	-0.320***	-0.033
	(0.074)	(0.075)	(0.119)	(0.139)
Mother's Edu Sq × Born 1968-72 × Inpres			0.023**	0.006
			(0.010)	(0.014)
R2	0.290	0.378	0.290	0.378
Observations	1083281	738202	1083281	738202

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Mother's Edu represents mother's years of schooling, which was calculated based on the education level completed. Family background is measured by mother's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table. The tables with all coefficients are available upon request. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table B.2: **Effects of Inpres Schools on Intergenerational Educational Mobility: Comparison Cohorts: 1958-1962**

Panel A: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.205***	0.215***	0.196***	0.199***
	(0.041)	(0.043)	(0.042)	(0.046)
Father's Edu × Born 1968-72 × Inpres	-0.022***	-0.014**	-0.039***	-0.019**
	(0.006)	(0.006)	(0.011)	(0.010)
Father's Edu Sq × Born 1968-72 × Inpres			0.002***	0.001**
			(0.001)	(0.001)
R2	0.105	0.170	0.116	0.186
Observations	1190828	837014	1190828	837014

Panel B: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	1.597***	0.551	1.534***	0.521
	(0.384)	(0.462)	(0.381)	(0.488)
Father's Edu × Born 1968-72 × Inpres	-0.123**	0.054	-0.234***	-0.014
	(0.050)	(0.061)	(0.085)	(0.092)
Father's Edu Sq × Born 1968-72 × Inpres			0.013**	0.008
			(0.006)	(0.007)
R2	0.321	0.408	0.322	0.408
Observations	1190828	837014	1190828	837014

Notes: Robust standard errors in parentheses clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1958 and 1962, or 1968 to 1972. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table B.3: **Effects of Inpres Schools on Intergenerational Educational Mobility: Comparison Cohorts: 1959-1962**

Panel A: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.210***	0.219***	0.199***	0.202***
	(0.041)	(0.041)	(0.043)	(0.044)
Father's Edu × Born 1968-72 × Inpres	-0.022***	-0.013**	-0.037***	-0.017*
	(0.006)	(0.006)	(0.012)	(0.010)
Father's Edu Sq × Born 1968-72 × Inpres			0.002***	0.001*
			(0.001)	(0.001)
R2	0.103	0.166	0.114	0.181
Observations	1178986	822791	1178986	822791

Panel B: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	1.666***	0.656	1.570***	0.609
	(0.389)	(0.467)	(0.396)	(0.494)
Father's Edu × Born 1968-72 × Inpres	-0.131**	0.066	-0.208**	0.017
	(0.051)	(0.061)	(0.097)	(0.091)
Father's Edu Sq × Born 1968-72 × Inpres			0.010	0.006
			(0.007)	(0.007)
R2	0.320	0.405	0.321	0.405
Observations	1178986	822791	1178986	822791

Notes: Robust standard errors in parentheses clustered at the district of birth(*** $p<0.01$, ** $p<0.05$, * $p<0.1$). Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1959 and 1962, or 1968 to 1972. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table; full table, with all coefficients, is available upon request. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table B.4: **Effects of Inpres Schools on Intergenerational Educational Mobility: Comparison Cohorts: 1960-1962**

Panel A: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	0.207***	0.203***	0.198***	0.190***
	(0.040)	(0.040)	(0.042)	(0.042)
Father's Edu × Born 1968-72 × Inpres	-0.022***	-0.012**	-0.038***	-0.017*
	(0.005)	(0.006)	(0.011)	(0.009)
Father's Edu Sq × Born 1968-72 × Inpres			0.002***	0.001**
			(0.001)	(0.001)
R2	0.100	0.159	0.111	0.174
Observations	1163653	804747	1163653	804747

Panel B: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1968-72 × Inpres	1.660***	0.551	1.602***	0.498
	(0.384)	(0.467)	(0.387)	(0.492)
Father's Edu × Born 1968-72 × Inpres	-0.131**	0.076	-0.247***	0.042
	(0.051)	(0.061)	(0.090)	(0.088)
Father's Edu Sq × Born 1968-72 × Inpres			0.014**	0.005
			(0.006)	(0.007)
R2	0.318	0.401	0.319	0.401
Observations	1163653	804747	1163653	804747

Notes: Robust standard errors in parentheses clustered at the district of birth (*** $p<0.01$, ** $p<0.05$, * $p<0.1$). Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1960 and 1962, or 1968 to 1972. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table; full table, with all coefficients, is available upon request. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table B.5: **Effects of Inpres Schools on Intergenerational Educational Mobility : Partially Exposed Cohorts Only**

Panel A: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1963-67 × Inpres	0.611**	-0.537*	0.569**	-0.491
	(0.270)	(0.323)	(0.286)	(0.335)
Father's Edu × Born 1963-67 × Inpres	-0.067	0.100*	-0.112	0.027
	(0.042)	(0.058)	(0.088)	(0.087)
Father's Edu Sq × Born 1963-67 × Inpres			0.006	0.007
			(0.006)	(0.007)
R2	0.317	0.403	0.317	0.403
Observations	433272	399906	433272	399906

Panel B: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born 1963-67 × Inpres	0.094***	0.040	0.082***	0.030
	(0.028)	(0.025)	(0.030)	(0.030)
Father's Edu × Born 1963-67 × Inpres	-0.013***	-0.003	-0.020**	-0.001
	(0.004)	(0.003)	(0.008)	(0.009)
Father's Edu Sq × Born 1963-67 × Inpres			0.001**	0.000
			(0.000)	(0.001)
R2	0.128	0.194	0.141	0.214
Observations	433272	399906	433272	399906

Notes: Robust standard errors in parentheses clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). Quadratic model is adopted following Becker et al. (2015). Sample corresponds to children born between 1957 to 1967. Covariates include birth district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. Father's Edu represents father's years of schooling, which was calculated based on the education level completed. Family background is measured by father's years of schooling. For the sake of parsimony, only intercept, linear, and quadratic terms are reported in this table; full table, with all coefficients, is available upon request. Data sources: Indonesia's full count census 2000 and Duflo (2001).

Table B.6: **Test of Truncation Bias: Evidence from IFLS and Census 2000**

	Linear CEF			Quadratic CEF		
	IFLS	**Census**	**IPW Corrected**	**IFLS**	**Census**	**IPW Corrected**
	(1)	(2)	(3)	(4)	(5)	(6)
			Daughters-Years of Schooling			
Father's Edu	0.522***	0.489***	0.511***	0.501***	0.480***	0.516***
	(0.021)	(0.007)	(0.007)	(0.053)	(0.013)	(0.014)
Father's Edu Sq.				0.002	0.001	-0.000
				(0.004)	(0.001)	(0.001)
Observations	2,362	848,350	848,350	2,362	848,350	848,350
R-squared	0.484	0.408	0.428	0.484	0.408	0.428
			Daughters-Primary Completion			
Father's Edu	0.023***	0.019***	0.025***	0.056***	0.043***	0.055***
	(0.002)	(0.001)	(0.001)	(0.005)	(0.002)	(0.002)
Father's Edu Sq.				-0.003***	-0.002***	-0.003***
				(0.000)	(0.000)	(0.000)
Observations	2,362	848,350	848,350	2,362	848,350	848,350
R-squared	0.309	0.160	0.204	0.325	0.177	0.225
			Sons-Years of Schooling			
Father's Edu	0.455***	0.433***	0.452***	0.571***	0.373***	0.408***
	(0.024)	(0.006)	(0.006)	(0.068)	(0.011)	(0.012)
Father's Edu Sq.				-0.008**	0.005***	0.004***
				(0.004)	(0.001)	(0.001)
Observations	2,361	1,199,814	1,199,814	2,361	1,199,814	1,199,814
R-squared	0.389	0.321	0.330	0.390	0.321	0.330
			Sons-Primary Completion			
Father's Edu	0.013***	0.013***	0.017***	0.035***	0.029***	0.037***
	(0.002)	(0.001)	(0.001)	(0.006)	(0.001)	(0.001)
Father's Edu Sq.				-0.002***	-0.001***	-0.002***
				(0.000)	(0.000)	(0.000)
Observations	2,361	1,199,814	1,199,814	2,361	1,199,814	1,199,814
R-squared	0.263	0.101	0.129	0.272	0.111	0.142

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Years of schooling (Edu) in Census 2000 was calculated based on the education level completed. Years of schooling (Edu) in the IFLS was calculated based on highest grade completed in an education level. Primary completion takes the value of 1 if a child has completed 6 or more years of schooling and 0 otherwise. Family background is measured by father's years of schooling. The Inverse Probability Weighting (IPW) estimates are calculated using 1975 district level density interacted with time trend. Data sources: Indonesia's full count census 2000, IFLS and IFLS-East.

Table B.7: **Effects of Inpres on Intergenerational Educational Mobility with Inverse Probability Weighting (IPW) Correction**

Panel A: Daughters

	Years of Schooling		Primary Completion	
	Linear CEF		Quadratic CEF	
	Unweighted (1)	IPW Corrected (2)	Unweighted (3)	IPW Corrected (4)
Born Between 1968-1972 × Inpres	0.513	0.432	0.197***	0.174***
	(0.466)	(0.431)	(0.047)	(0.052)
Father's Edu × Born 1968-72 × Inpres	0.043	0.021	-0.017*	-0.015
	(0.062)	(0.057)	(0.010)	(0.013)
Father's Edu Sq × Born 1968-72 × Inpres			0.001*	0.001
			(0.001)	(0.001)
Observations	848,350	848,350	848,350	848,350
R-squared	0.409	0.429	0.188	0.240

Panel B: Sons

	Years of Schooling		Primary Completion	
	Quadratic CEF		Quadratic CEF	
	Unweighted (1)	IPW Corrected (2)	Unweighted (3)	IPW Corrected (4)
Born 1968-72 × Inpres	1.534***	1.553***	0.195***	0.201***
	(0.380)	(0.376)	(0.042)	(0.040)
Father's Edu × Born 1968-72 × Inpres	-0.255***	-0.293***	-0.038***	-0.033***
	(0.086)	(0.097)	(0.011)	(0.012)
Father's Edu Sq × Born Between 1968-72 × Inpres	0.015**	0.018**	0.002***	0.002***
	(0.007)	(0.007)	(0.001)	(0.001)
Observations	1,199,814	1,199,814	1,199,814	1,199,814
R-squared	0.323	0.332	0.117	0.153

Notes: Robust standard errors in parentheses clustered at the district of birth (*** p<0.01, ** p<0.05, * p<0.1). The CEFs are based on Table A.2. Quadratic model is adopted following Becker et al. (2015). Unweighted estimates are same as estimates reported in Table 2 and Table 3 for years of schooling and primary completion outcomes, respectively. Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Years of schooling (Edu) in Census 2000 was calculated based on the education level completed. Years of schooling (Edu) in the IFLS was calculated based on highest grade completed in an education level. Primary completion takes the value of 1 if a child has completed 6 or more years of schooling and 0 otherwise. Family background is measured by father's years of schooling. The Inverse Probability Weighting (IPW) estimates are calculated using 1975 district level density interacted with time trend. Data sources: Indonesia's full count census 2000, Duflo (2001), IFLS and IFLS-East.

Table B.8: **Effects of Inpres on Coresidency Rates**

	Sons	**Daughters**
	(1)	(2)
Born 1968-72 $\times$ Inpres	-0.036	-0.036
	(0.022)	(0.023)
Constant	0.520***	0.517***
	(0.005)	(0.005)
R2	0.042	0.058
Observations	1057100	720747

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1) Sample corresponds to children born between 1957 and 1962, or 1968 to 1972. Covariates include birth district district FE, year of birth×1971 enrollment, year of birth×1971 number of children, year of birth×water sanitation program, year of birth dummies, following Duflo (2001). The variable Inpres measures the number of Inpres schools per 1000 children at the district level divided by the highest number of schools received by one district. The census 2000 collected data on number of children born alive to an adult woman. The coresidency rate is the ratio of number of children in the household divided by number of children born alive. Data sources: Indonesia's full count census 2000 and Duflo (2001).