Unintended bottleneck and essential nonlinearity: Understanding the effects of public primary school expansion on intergenerational educational mobility

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

Saint Louis University, World Bank, IPD, Columbia

9 May 2022

Unintended Bottleneck and Essential Nonlinearity: Understanding the Effects of Public Primary School Expansion on Intergenerational Educational Mobility

Md. Nazmul Ahsan, St. Louis University[1]

M. Shahe Emran, IPD at Columbia University

Forhad Shilpi, DECRG, World Bank

First Draft: April 17, 2022
This version: May 9, 2022

Comments Welcome

ABSTRACT

We study the effects of 61,000 public primary schools on intergenerational educational mobility in Indonesia using full-count census data, a credible identification strategy, and theory based nonlinearity in the mobility equation. We find that the mobility curve is concave in most of the cases, and school expansion reduced the degree of concavity. Evidence from a DiD strategy (Duflo (2001)) on primary completion suggests substantial improvements in relative mobility of the children of low educated fathers irrespective of gender. But relative mobility in the college educated households worsened, strengthening the advantages of the better educated households across generations. This highlights the pitfalls of a linear model which incorrectly suggests a weakening of the advantages of the children of educated fathers. For completed years of schooling, there are striking gender differences: the strong effects on sons remain largely unchanged, but there are no significant effects on girls. The surprising absence of an effect on girls is due to an unintended bottleneck at the secondary schooling level creating fierce competition among the Inpres primary graduates. The girls lost ground, experiencing 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 girls suffered crowding out irrespective of the family background, suggesting that social norms rather than parental economic conditions are the mechanisms at work.

Key Words: Public Schools, Intergenerational Mobility, Education, Theory-based Nonlinearity, Indonesia, Pitfalls of Linearity, Gender Bias, Social Norms, Big Data

JEL Codes: I24, J62, J16, O20

[1]Emails for correspondence: ahsanm@slu.edu (Md Nazmul Ahsan) , shahe.emran.econ@gmail.com (M. Shahe Emran), fshilpi@worldbank.org (Forhad Shilpi). We would like to thank Shailesh Tiwari, Imam Setiawan, and Rinku Murgai for help with data. 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 50 years.[2] A widely espoused policy rationale for such school expansion is that a better access to schools helps reduce inequality by improving educational mobility of the children from disadvantaged socioeconomic background. However, the effects of school expansion on intergenerational educational mobility remains largely unexplored in the recent economic literature on developing countries. While the construction of a public tuition-free primary school in a village seems to fit the description of canonical government policies for expanding access to education for all, the incidence of such policies may not be distributionally progressive or neutral.[3] It has long been recognized that the effects of government policies on children may vary depending on the family background and gender of a child (Becker (1981), Atkinson and Stiglitz (1980)). 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. Gender of a child may be important as distance to schools hinders the girls more because parents are unwilling to send them to far away schools for fear of harassment on the way to school, for example (World Bank (2018), Tilak (1993), Scott (1985)).

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. 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. The focus of our analysis is on two issues: (1) potential non-linearity in the intergenerational mobility equation, 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, with the standard linear specification as a special case. The concavity arises naturally

[2]For a comprehensive discussion on the evolution of school access in developing countries, see chapter 2 titled "The great school expansion- and those it has left behind" in World Bank (2018).

[3]The "free" public schools may not be free, especially for poor children, when parents need to pay bribes for admission into public schools. In Bangladesh, about 50 percent of parents end up paying bribes for admission of their children into "free" public schools. More important, the unfortunate half that pay bribes comprises of mainly the poor parents as they have little bargaining power to resist such bribe demands, while the rich get their children admitted without paying bribes. For details, see Emran et al. (2020).

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. An important implication of a concave or convex mobility equation is that relative mobility (as measured by intergenerational regression coefficient) is not constant across the distribution of father's schooling, unlike the workhorse linear model. The expansion of private schools can affect the shape of nonlinearity, implying substantially different effects on the relative mobility of the children from different parts of the father's schooling distribution.[4]

The second question we focus on is whether the primary school expansion had differential effects across boys and girls. As noted above, there is a broad consensus that proximity to schools matters much more for the girls as parents are in general unwilling to send daughters to far away schools because of safety concerns. If distance to school is a binding constraint on girls schooling, then we would expect a bigger impact of the Inpres schools on the educational opportunities of daughters.

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). We are not aware of any other analysis of intergenerational educational mobility in a developing country that takes advantage of such a large data set.[5] 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 2.2 million father-child pairs.[6] We provide evidence that the estimates from the census data do

[4]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. An alternative approach for analyzing nonlinearity is completely data driven: estimating nonparametrically the relation between parent's and children's economic status without any explicit theoretical basis for the observed nonlinearity.

[5]The influential contributions by Chetty et al. (2014) have highlighted the advantages of big data in understanding intergenerational economic mobility. Card et al. (2018) use 1940 census data to analyze intergenerational educational mobility in USA at the beginning of 20th century. A number of recent papers on intergenerational educational mobility in Africa uses the 10 percent sample of the census available through IPUMS; see, for example, Alesina et al. (2021) and Azomahou and Yitbarek (2021).

[6]Such a large sample size ensures that when we fail to reject a zero null hypothesis, it is not because of large standard errors in a small sample.

not suffer from any significant bias relative to a widely used data set for estimating intergenerational effects in Indonesia: the Indonesia Family Life Survey (IFLS). We, however, do not use the IFLS data for our empirical analysis because of the small sample size given the cohorts restrictions in our analysis.[7]

The evidence suggests four 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 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 estimates thus suggest that the children from the most disadvantaged background (father's with no schooling) enjoyed higher relative mobility (negative effect on the linear term) and absolute mobility (positive effect on the intercept). Second, the estimates from the widely used linear model can be misleading, especially for understanding relative mobility because, by assumption, the impact on relative mobility does not vary with the family background of a child in a linear CEF. The estimates from the correct concave model show important heterogeneity: while the Inpres schools improved relative mobility of the children from low educated households, it had the opposite effect 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. The standard linear model also underestimates substantially the improvements in relative mobility experienced by the most disadvantaged children born to fathers with no schooling (72 percent underestimation for sons and 22 percent for daughters). 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 important, the evidence suggests strong effects of Inpres schools on sons, but there are no significant effects on daughters.[8] Fourth, we provide an explanation for the puzzle of strong effects on primary schooling of girls, but

[7]The seminal analysis of the effects of Inpres schools in Indonesia by Duflo (2001) uses the Inter-Censal Population Survey (SUPAS 1995) data and thus is not subject to the small sample problem.

[8]It is striking that even with such a large sample size (848,350 observations) the effects on daughters are statistically not significant at the 10 percent level. The estimated effects are also much smaller in magnitude.

little effects on the completed years of schooling. The 61k 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 suggests that the boys experienced a positive effect (7.7 percentage points higher probability of completion), but the girls experienced a negative effect (8.5 percentage points lower probability of completion). The evidence thus suggests that the expanded supply of primary graduate boys crowded out the girls from the high school.[9] The Inpres schools had an unintended negative effect on gender equality in higher education of the exposed cohorts. The evidence suggests that the negative effect on a girl's secondary schooling does not depend on her socioeconomic background which is consistent with a primary role played by gender-based social norms rather than economic conditions of parents.[10]

The analysis and conclusions of this paper have wider implications. The evidence that the standard linear model might be seriously misleading, especially for understanding the effects of government policies on relative mobility, is important because linearity is a maintained assumption in many existing studies on the effects of government policies on income and educational mobility (see Lou and Li (2022), Pekkarinen et al. (2009), Parman (2011)).[11] This underscores the importance of using nonlinear mobility models based on theory as suggested by Cholli and Durlauf (2022). The finding that the schooling expansion at the primary level creates 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 level has lagged far behind.[12] The insight about possible unintended bottleneck is of more general relevance: any

[9]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.

[10]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)).

[11]The last two studies analyze the effects of school reform or expansion on intergenerational income mobility. The income mobility equation is linear in log, and thus relative mobility is estimated as elasticity (called IGE in the literature). What is important here is that relative mobility and the effects of policy on relative mobility do not vary across father's economic status in these empirical models.

[12]In 2016, the secondary completion rate was only 35 percent in the low income countries as classified by

government policy that expands 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 or religious minority, lower caste, girls).

The remainder of the paper is organized as follows. The next section 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. The next section reports the main empirical results including estimates of the effects of Inpres schools on relative and absolute mobility. Section (6) offers some 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 (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. The paper ends with a summary of the main findings and their implications for the broader literature on intergenerational mobility.

(2) Related Literature

The contributions of this paper relate to a number of active areas of economic research. The first and most obvious strand of related literature is intergenerational mobility, especially in the context of developing countries. 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), and excellent surveys by 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 Azam and Bhatt (2015), Alesina et al. (2021), Emran and Shilpi (2015), and Neidhofer et al. (2018).[13] Given the data

World Bank. The corresponding rate for OECD countries was 96 percent. See chapter 2 in World Bank (2018).

[13] Among unpublished papers, see Yu et al. (2020), Asher et al. (2018), Emran et al. (2021).

constraints, there are only a few papers that analyze income or expenditure persistence in developing countries: see Fan et al. (2020) on income mobility China and Dartanto et al. (2021) on expenditure mobility in Indonesia. 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 (see Orazem and King (2008), Hanushek (2002), and Filmer (2007)). As noted earlier, following the influential work of Duflo (2001, 2004), many papers have studied the effects of the Inpres schools in Indonesia; for example, Ashraf et al. (2020) provide evidence showing 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 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 in the first generation exposed to Inpres program on the school performance and other outcomes of the second generation of children. However, they do not study whether the effects on the first generation mothers depend on their socioeconomic background. If the Inpres schools improved girl's 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 in the first generation, we provide the critical missing link in understanding the long-term distributional consequences of the 61 k primary schools built under the Inpres program.

In an unpublished paper, Hertz and Jayasundera (2007) provide an analysis of the effects of Inpres schools on intergenerational educational mobility using IFLS 2000 round.[14] However, their results are difficult to interpret because of specification errors in their empirical models where they incorrectly exclude (son's specification) or include (daughter's specification) terms

[14] They are aware of potential power problems due to small sample size of the IFLS data, and discuss that their sample size is much smaller, about 3.4 percent of the sample size of Duflo's (2001) 1995 SUPAS data.

involving father's education squared (see footnote 21 in section 4 below for details). Also, they use an Inpres intensity measure which is different from that used by Duflo (2001) and all the subsequent papers noted above, including ours. This makes their estimates not comparable to ours. Another important difference from our analysis is that they focus exclusively on years of schooling as a measure of children's education, while we analyze the effects on primary and secondary completion in addition to years of schooling. As we show below, to understand the effects of Inpres schools on intergenerational mobility one needs to look at primary and secondary completion of children in addition to years of schooling. 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 theoretically grounded empirical specification that accounts for potential nonlinearity, full count census data, and a credible identification strategy. We are not aware of any other study that provides evidence on the effects of public school expansion on intergenerational educational mobility in a developing country.

In a recent paper, Card et al. (2018) provide an interesting analysis of the effects of public school expansions in the early 20th century USA using 1940 census data. Their focus is on the 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 low quality teachers because of salary caps affected the black children's educational attainment adversely. There are a few papers that look at the causal effects of school reform or expansion 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 on the effects of better school quality on intergenerational income mobility in Iowa. Parman (2011) reports evidence that improvements in school quality lowered 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 individuals who are 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. Therefore, 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. 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 education 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.

Given that we have the full count data, our main estimation sample (exposed: 1968-1972 and unexposed: 1957-1962) gives us about 2.2 million father-child pairs where household heads are the fathers, and the children were living at the household at the time of the census. The mother-children sample has about 1.8 million observations. We calculate the years of schooling based on the education level a respondent has completed. A comparison shows that the exposed children and their fathers have more education than comparison cohorts and their fathers (see Table 1). The sample for the partially exposed group consists of 628,343 father-child pairs. Again, the exposed group on average has more schooling.

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.[15]

We check whether the census estimates are substantially biased downward compared to the widely used data from Indonesia Family Life Survey (IFLS). Recent evidence shows that sample truncation due to nonrandom missing of children biases the estimated slope parameter downward in a linear educational mobility model, see Emran et al. (2018) and Azam and Bhatt (2015).[16] To understand whether the estimates from the census data are substantially biased

[15]The district that received the highest number of schools received about 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 almost zero (0.0678) to one with a mean of 0.214.

[16]We are not aware of any evidence on how sample truncation affects the parameters of a quadratic mobility

due to such missing children from a household, we take advantage of the fact that the IFLS survey collected information on the noresident children. If the estimates from census data are substantially biased the estimated slope parameter in a linear regression will be much smaller for the census data when compared to the IFLS data. Table A.1 in the online appendix presents the estimates of the linear mobility model for the census and IFLS data (including IFLS-East). The evidence is reassuring: the estimate of the slope parameter is smaller in census (compared to that from IFLS), but the differences in the magnitudes are small. For sons sample, the estimated slope is -0.49 in census data and -0.52 in IFLS data. For daughters, the corresponding estimates are -0.59 (census) and -0.62 (IFLS). The evidence thus suggests that the estimates from the census data suffer downward bias, but the degree of bias is not substantial enough to raise concerns about the validity of our analysis.

(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 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.

model.

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

In the online appendix, we provide additional evidence supporting the identification assumption by estimating a series of placebo regressions in the pre-Inpres periods. For examples of such placebos, please see online appendix Tables A.2 and A.3. The large sample size allows us to consider many "fictitious" Inpres dates for a variety of subsamples in the pre-Inpres period. A concern here is that the cohorts that were born much earlier than the actual Inpres implementation date may not be comparable to the Inpres cohorts under study, and thus can lead to misleading conclusions about the validity of the identifying assumptions. We thus report placebos by excluding the oldest cohorts in the data set. We also check if a fictitious Inpres date splitting our main comparison cohorts (1957-1962) into "exposed" and

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

‘unexposed” groups shows any evidence that contradicts the identifying assumption. Evidence clearly supports the identification strategy.

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:

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

where E_{ikt} is an indicator of educational attainment of child i, k denotes the birth district of child i, t denotes time period (year), superscripts c and p refer to children and parents respectively, 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. Inp_k is a measure of the intensity of the new school construction in district k. We use a normalized measure so that $Inp_k \in [0, 1]$.[18] 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.

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 $W_k\tau_t$): 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

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

for the linear model becomes:

$$E_{ikt}^{c} = \beta_0 + \beta_1 E_{ikt}^{p} + \beta_4 \left(Exp_t \times Inp_k\right) + \beta_5 \left(E_{ikt}^{p} \times Inp_k\right) \\ + \beta_6 \left(E_{ikt}^{p} \times Exp_t\right) + \beta_7 \left(E_{ikt}^{p} \times Exp_t \times Inp_k\right) + \alpha_k + \tau_t + \sum_t W_k \tau_t + \varepsilon_i \quad (2)$$

The focus of a large literature on intergenerational educational mobility is on the parameter β_1, called intergenerational regression coefficient (IGRC, for short) which is a measure of relative mobility (see, for example, Hertz et al. (2008), 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 (a larger estimated β_1) implies lower mobility.[19] We underscore an often overlooked limitation of the linear model: relative mobility is constant across the distribution by assumption, with the children of fathers with no schooling facing the same relative mobility as the children of college or more educated fathers. We also discuss the effects on the intercept 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. 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.

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 analysis 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

[19] Some authors use $1 - \beta_1$ as a measure of relative mobility.

relative mobility and the effects of policy interventions vary across the distribution of father's education, see Card et al. (2018), Asher et al. (2018), Emran et al. (2021), and Ahsan et al. (2021).

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 education 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.[20] 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):

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

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).[21] In a quadratic mobility model, the impact on

[20] Ahsan et al. (2021) also suggest that assortative matching on education in the marriage market can make the mobility CEF convex when only one parent's (usually father) education is used as a measure of family background.

[21] Comparing equation (3) with the specification used by Hertz and Jayasundera (2007) (see their equation (1) in P.8) suggests that their specification is misspecified. Hertz and Jayasundera (2007) include an interaction of parent's education squared with exposure to Inpres schools but *incorrectly exclude parent's education squared itself from the empirical model.* This exclusion inadvertently imposes a maintained assumption that the CEF

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 father's schooling. The estimate of θ_7 is the effects on relative mobility of the children of father's 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}.

In a quadratic mobility model, a natural extension of the standard measure of relative mobility 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 with that of the parents.[22] Relative mobility for the children of fathers with y years of schooling is given as:

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

(5) Empirical Evidence

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

was linear in the pre-Inpres period which is rejected by data for sons, as our results for years of schooling show below. The CEF for the daughter's years of schooling, on the other hand, is linear in both pre-Inpres and Inpres cohorts (see the evidence in online appendix Table A.2). This suggests that the interaction of father's education squared with Inpres exposure in their specification should be excluded in the case of daughters. It is thus not clear how to interpret their estimates.

[22] 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.

large literature on intergenerational educational mobility (see the surveys by Torche (2019), Iversen et al. (2019), and Emran and Shilpi (2021)). If we use mother's education in place of father's education, the sample size is smaller, but the main conclusions of the paper are not affected (see section 6 below for the evidence and discussion). All reported standard errors are clustered at the district level.

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

In this subsection we discuss the empirical results 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 panel A of Table 2 report the estimates from the linear model, while the last two report the estimates from the quadratic model. Table 2 reports the estimated effects on the parameters of interest (intercept, linear and quadratic terms of the mobility models), and the full set of coefficient estimates for the corresponding DiD models are reported in online appendix Table A.4.

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) $\left(\hat{\beta}_7^s = -0.13\right)$.[23] 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 father's education level. The positive effect on the intercept suggests that the most disadvantaged children 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 has failed to affect the 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. It is important to underscore that the insignificant effect on girls is not because of sample size: the estimates are based on a sample of 848,350 observations.[24]

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

[24] In an unpublished working paper, Hertz and Jayasundera (2007) also report no significant effects for girl's years of schooling, but do not analyze the effects on primary or secondary schooling. More important,

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.5 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. 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 quadratic model (estimating equation (3)) and the estimated full specification is provided in the online appendix Table A.4. 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.[25] 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

it is difficult to be confident about their finding because, as noted earlier, their empirical specification is misspecified. Another worry is the relatively small sample size in their estimation because they use IFLS 2000 round.

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

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 panel B of Table 2 (the full set of coefficients are reported in online appendix Table A.4).

We first consider the estimates from the linear model (estimating equation (2)). 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). The point estimates for sons (superscript s) and daughters (superscript d) are: Intercept $\hat{\beta}_4^s = 0.20$ and $\hat{\beta}_4^d = 0.22$, and slope $\hat{\beta}_7^s = -0.022$ and $\hat{\beta}_7^d = -0.014$.[26] This strong effect is more consistent with the a priori expectation that distance to school is more of a constraint for girls because of safety concerns, among other things. 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 Table A.5 in the online appendix. The evidence suggests that the 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 a quadratic model (as in equation (3)) would be more appropriate for both sons and daughters to understand the effects of Inpres schools on primary school completion.

The estimates of the relevant parameters from the quadratic model are reported in columns (3) (sons) and (4) (daughters) in the lower panel of Table 2. The pattern of the estimated

[26] For daughters, the intercept effect is significant at the 1 percent level, and the slope effect at the 5 percent level.

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 $\left(\hat{\theta}_{11}^s = 0.002; \hat{\theta}_{11}^d = 0.001\right)$.[27] 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 (junior secondary) is 0.038 which equals the the linear coefficient in magnitude $\left(\hat{\theta}_7^s = -0.038\right)$. Similar conclusions hold for the daughter's estimates. We provide estimates of changes in relative and absolute mobility due to the advent of the Inpres schools in the next subsection below.

(5.3) 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. We also discuss the effects on absolute mobility as measured by expected educational outcome of a child conditional on his/her father's education.

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.6 for the effects on son's 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.

As noted earlier, our measure of relative mobility is IGMA (intergenerational marginal association) which is an extension of the standard measure in the linear case called IGRC

[27] 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.

and is estimated as the slope of the mobility CEF. Absolute mobility is measured by the expected years of schooling conditional on father's education which is given by the point on the estimated CEF corresponding to the level of father's education. This measure follows the recent influential work of Chetty et al. (2014). 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.[28]

Effects on Relative Mobility

Using the estimated coefficients in Table 2, 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 is give by:

$$\Delta IGMA_{iy} = \theta_7 + 2\theta_{11} E^p_{iy}$$

We provide estimates for two levels of treatment intensity: the mean level and the highest intensity in our data.[29]

The evidence in Table 3 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^s_0 (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, and by 0.82 percentage points in a district of 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 education) in the zero Inpres intensity districts in the pre-Inpres period as the benchmark.[30] The normalized

[28] In the special case of a linear CEF, this also gives the expected income rank for the children from the lowest half of parental distribution. Some authors use parent's education as a reference point, defining upward absolute mobility when children have more schooling than their parents. Cardona and Jones (2021) combines the CEF based measure with father as a reference point.

[29] The mean is 1.86 new schools per 1000 school age children.

[30] Recall that IGMA is the slope of the mobility CEF. In Table 3, IGMA gives the change in the probability of completing primary schooling when father's education increases by one year. The estimates reported in Table 3 are *changes in IGMA* caused by Inpres schools. The corresponding estimates using an alternative benchmark, the unexposed children *born to fathers with no schooling* in districts with zero Inpres intensity,

estimates relative to the benchmark are reported in the odd numbered columns in Table 3. For the sons born to fathers with no schooling, the Inpres schools reduced the IGMA by 31.18 percent in a district with mean Inpres intensity, and by 145 percent in a district with the highest treatment intensity (relative to the IGMA of the benchmark described above). For daughters in this subgroup (fathers with no schooling), the corresponding improvements in relative mobility (reductions in IGMA_0^d) are smaller in magnitude: 13.95 percent (average intensity) and 64.89 percent (highest intensity).

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 21.34 percent higher in a district with average treatment intensity, and 99.24 percent higher in a district with the highest Inpres intensity. Similar conclusions hold for daughters. This suggests that the inheritance of educational status became much more persistent across generations at the top of the education distribution.

Effects on Absolute Mobility

As noted earlier, absolute mobility is measured by the expected educational outcome of a child conditional on father's education. The estimated effects of Inpres schools on the expected probability of having primary or more schooling, and on expected years of schooling of children are reported in the lower panel of Table 3 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 zero Inpres intensity.

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 (father having no schooling), there is no gender gap: both the sons and daughters experienced a 20 percent higher probability of having primary or more schooling in districts with the highest treatment intensity, about 4.5 percent higher probability at the districts with average

are in online appendix Table A.6.

Inpres intensity (relative to the expected years of schooling of the benchmark group). In contrast, there is clear gender differences in the college educated households: the daughters benefited almost twice as much as the sons: the daughters reaping a 19 percent higher probability of having primary or more education while the sons gaining about 10.4 percent higher probability.

(6) Robustness Checks: Mother's Education, Alternative Unexposed Groups, and Partial Exposure

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 could 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 4.[31] The evidence is very similar to what we found earlier in Table 2 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, there are substantial and statistically significant effects at the primary level.

The DiD design in Table 2 uses data on 6 years from the pre-Inpres period to define the unexposed 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 unexposed sample. We report the estimates for these alternative samples in online appendix Tables A.8-A.10. The results are robust across these alternative pre-Inpres cohorts.

[31] The full DiD model estimates are in online appendix Table A.7.

We also provide estimated effects for the partially exposed cohorts (1963-1967) in the online appendix: please see Table A.11. The estimated effects are numerically much smaller as one would expect, and many are not significant at the 10 percent level. We also provide the estimates that combine the partially exposed cohorts with our main exposed cohorts which shows evidence that inclusion of the partially exposed group dilutes the estimated effects substantially; see Table A.12 in the online appendix.

(7) The Consequences of an Unintended Bottleneck: An Explanation for the 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 mechanism can give rise to this gender difference.

The construction of 61k 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 as to 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 limited number of high schools. If this is the case we will see that 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.[32] A testable implication of this hypothesis is that the impact of father's education at the high school level should become much stronger. This mechanism on its own, however, 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.[33] If cost of safe transport to far away schools is the binding constraint,

[32]One would expect bribery and "donations" to play important role in who gets admitted.

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

then we would still expect substantial effects of family background, where only the girls from low educated (and low income) households fail to progress beyond primary schooling. But an alternative gender specific hypothesis is that in the 1970s and early 1980s, most of the parents considered primary completion as a socially expected norm for girls, and paid little attention to their progression beyond this level. If this is the case, then we should not see any significant difference across the daughters of more and less educated fathers.

To explore these questions, we estimate the effects of Inpres schools on two binary indicators of secondary schooling attainment by children: junior high (3 years after 6 years primary) and more (henceforth "junior high"), and senior high (3 years after junior high) and more (henceforth "senior high") schooling of children. Note that the linear model is rejected for both the junior and senior high as measures of educational attainment of children. Please see online appendix Table A.11 for details. The estimates from the quadratic model are reported in Table 4.

The first striking finding from Table 4 is that family background (as measured by father's education) plays virtually no role in determining the impact of Inpres school expansion on secondary schooling: the triple interactions 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). This is true for both junior and senior high school results, and holds irrespective of gender. The estimated effects of Inpres on the intercepts suggest a different picture: it is negative and statistically significant for girls at both junior high and senior high levels. The estimate is much larger at the junior high level implying that the girls in a district with the highest treatment intensity suffered a 15.7 percentage points decline in the probability of completing junior secondary high schooling irrespective of socioeconomic background. The negative effect is smaller at the senior secondary level, but still substantial: a 8.5 percentage points lower probability of having senior high schooling for the Inpres cohorts. The evidence for the sons in contrast suggests that they have gained, especially at the senior high level: 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. The evidence that the gender-based crowding out happened across the board suggests that it is primarily due to social norm unrelated to the socioeconomic background of a child.

(8) **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 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.

Evidence rejects the standard linear model in favor of a concave mobility model in all our empirical models except for the case when years of schooling is used as the measure of educational attainment of girls. We find that Inpres schools made the mobility CEF less concave (or more convex), again except for the case of girl's years of schooling with no significant effect on the shape of the mobility curve. This led to substantial improvements in relative mobility of the disadvantaged children 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 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 father's education level. This supports the hypothesis that the primary mechanism behind the gender-based crowing out is social norms against girls, rather than differences in the economic conditions between more and less educated households.

References

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.

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.

Asher, S., Novosad, P., and Rafkin, C. (2018). Intergenerational Mobility in India: Estimates from New Methods and Administrative Data. Working paper, World Bank.

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

Atkinson, A. B. and Stiglitz, J. E. (1980). *Lectures on Public Economics Updated edition*. McGraw Hill Books.

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.

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.

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.

Card, D., Domnisoru, C., and Taylor, L. (2018). The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility. NBER Working Papers 25000, National Bureau of Economic Research, Inc.

Cardona, M. and Jones, S. (2021). How far does the apple really fall from the tree?: Practical guidance on measuring intergenerational mobility from a simulation framework. WIDER Working Paper Series wp-2021-101, World Institute for Development Economic Research (UNU-WIDER).

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.

Dartanto, T., Can, C. K., Moeis, F. R., and Rezki, J. F. (2021). A dream of offspring: Two decades of intergenerational economic mobility in indonesia. *Bulletin of Indonesian Economic Studies*, 0(ja):1–50.

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., 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., Islam, A., and Shilpi, F. (2020). Distributional Effects of Corruption When Enforcement is Biased: Theory and Evidence from Bribery in Schools in Bangladesh. *Economica*, 87(348):985–1015.

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.*

Fan, Y., Yi, J., and Zhang, J. (2020). Rising Intergenerational Income Persistence in China. *American Economic Journal: Economic Policy.*

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. and Jayasundera, T. (2007). School Construction and Intergenerational Mobility in Indonesia. Working Papers 2007-18, American University, Department of Economics.

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.

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.

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

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.

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.

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

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.

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. (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.

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).

Tables

Table 1: Summary statistics

Panel A: Child–Father Sample		
	Mean	SD
Exposed Cohort (Born Between 1968-1972), Obs= 1,959,977		
Child's Years of Schooling	9.36	4.21
Father's Years of Schooling	5.73	4.63
Comparison Cohort (Born Between 1957 to 1962), Obs= 268,622		
Child's Years of Schooling	7.01	4.73
Father's Years of Schooling	4.43	4.36
Panel B: Child–Mother Sample		
Exposed Cohort (Born Between 1968-1972), Obs= 1,771,489		
Child's Years of Schooling	9.43	4.19
Mother's Years of Schooling	4.93	4.18
Comparison Cohort (Born Between 1957 to 1962), Obs= 212,373		
Child's Years of Schooling	7.13	4.73
Mother's Years of Schooling	3.71	3.97

Notes: Data source: BPS Census 2000

Table 2: Effects of Inpres Schools on Intergenerational Educational Mobility–Father's Years of Schooling as Family Background

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	1.585***	0.513	1.534***	0.476
	(0.387)	(0.466)	(0.380)	(0.492)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.129**	0.043	-0.255***	-0.020
	(0.054)	(0.062)	(0.086)	(0.099)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.015**	0.008
			(0.007)	(0.008)
R2	0.322	0.409	0.323	0.409
Observations	1199814	848350	1199814	848350

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.204***	0.216***	0.195***	0.197***
	(0.041)	(0.044)	(0.042)	(0.047)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.022***	-0.014**	-0.038***	-0.017*
	(0.006)	(0.006)	(0.011)	(0.010)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.002***	0.001*
			(0.001)	(0.001)
R2	0.107	0.172	0.117	0.188
Observations	1200144	848619	1200144	848619

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to individuals 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. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Years of schooling was calculated based on the education level completed. 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 A4. Data sources: BPS Census 2000 and Duflo (2001).

Table 3: Effects of Inpres Schools on Relative Mobility and Absolute Mobility in Primary Completion

Panel A: Intergenerational Marginal Association (IGMA) / Relative Mobility (Sons)				
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
	△IGMA	**Normalized △IGMA**	**△IGMA**	**Normalized △IGMA**
$IGMA_0$	-0.0380	-145.04%	-0.0082	-31.18%
$IGMA_6$	-0.0140	-53.44%	-0.0030	-11.49%
$IGMA_9$	-0.0020	-7.63%	-0.0004	-1.64%
$IGMA_{12}$	0.0100	38.17%	0.0022	8.21%
$IGMA_{16}$	0.0260	99.24%	0.0056	21.34%
Panel B: Intergenerational Marginal Association (IGMA) / Relative Mobility (Daughters)				
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
	△IGMA	**Normalized △IGMA**	**△IGMA**	**Normalized △IGMA**
$IGMA_0$	-0.0170	-64.89%	-0.0037	-13.95%
$IGMA_6$	-0.0050	-19.08%	-0.0011	-4.10%
$IGMA_9$	0.0010	3.82%	0.0002	0.82%
$IGMA_{12}$	0.0070	26.72%	0.0015	5.74%
$IGMA_{16}$	0.0150	57.25%	0.0032	12.31%
Panel C: Expected Schooling (ES)/ Absolute Mobility (Sons)				
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
	△ES	**Normalized △ES**	**△ES**	**Normalized △ES**
ES_0	0.1950	20.49%	0.0419	4.40%
ES_6	0.0390	4.10%	0.0084	0.88%
ES_9	0.0150	1.58%	0.0032	0.34%
ES_{12}	0.0270	2.84%	0.0058	0.61%
ES_{16}	0.0990	10.40%	0.0213	2.24%
Panel D: Expected Schooling (ES) / Absolute Mobility (Daughters)				
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
	△ES	**Normalized △ES**	**△ES**	**Normalized △ES**
ES_0	0.1970	20.70%	0.0424	4.45%
ES_6	0.1310	13.76%	0.0282	2.96%
ES_9	0.1250	13.13%	0.0269	2.82%
ES_{12}	0.1370	14.39%	0.0295	3.09%
ES_{16}	0.1810	19.01%	0.0389	4.09%

Notes:$\Delta IGMA$ represent the effects of the Inpres schools on the change in relative mobility. ΔES represent the effects of the Inpres schools on the change in absolute mobility. 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. $IGMA_k$ and ES_k are calculated at k father's years of education, where $k = 0, 6, 9, 12, 16$. The IGMA and ES values are based on coefficients reported in Panel B of Table A4. The Normalized IGMA/ES is the IGMA/ES value relative to IGMA/ES of children for father's with average years of education (5.57 years), unexposed cohort (born between 1957 to 1962), and zero Inpres intensity for the full sample (combined sample of sons and daughters). Normalized IGMA and ES values are reported in percentage. Data sources: BPS Census 2000 and Duflo (2001).

Table 4: Effects of Inpres Schools on Intergenerational Educational Mobility–Mother's Years of Schooling as Family Background

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	1.603***	0.755	1.592***	0.720
	(0.401)	(0.462)	(0.396)	(0.487)
Mother's Years of Schooling × Born Between 1968-1972 × Inpres	-0.129*	0.010	-0.320***	-0.033
	(0.074)	(0.075)	(0.119)	(0.139)
Mother's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.023**	0.006
			(0.010)	(0.014)
R2	0.290	0.378	0.290	0.378
Observations	1083281	738202	1083281	738202

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.184***	0.225***	0.180***	0.212***
	(0.044)	(0.043)	(0.044)	(0.044)
Mother's Years of Schooling × Born Between 1968-1972 × Inpres	-0.021**	-0.020***	-0.041***	-0.023*
	(0.008)	(0.007)	(0.014)	(0.012)
Mother's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.003***	0.001
			(0.001)	(0.001)
R2	0.099	0.157	0.107	0.169
Observations	1083576	738428	1083576	738428

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Sample corresponds to individuals 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. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Years of schooling was calculated based on the education level completed. 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: BPS Census 2000 and Duflo (2001).

Table 5: Effects of Inpres Schools on Intergenerational Educational Mobility–Junior and Senior High School Completion as Measures of Children's Educational Attainment

	Junior High Completion		**Senior High Completion**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.021	-0.157**	0.077*	-0.085*
	(0.053)	(0.073)	(0.039)	(0.049)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	0.001	0.029	-0.007	0.009
	(0.011)	(0.018)	(0.011)	(0.011)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres	-0.000	-0.001	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
R2	0.239	0.328	0.270	0.343
Observations	1200144	848619	1200144	848619

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Sample corresponds to individuals 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. Junior high completion takes the value of 1 if the child has completed junior high and 0, otherwise. Senior high completion takes the value of 1 if the child has completed senior high 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. 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: BPS Census 2000 and Duflo (2001).

Appendix Tables

Table A1: Linear CEF Comparison from IFLS and Census 2000

	IFLS Sample		**Census 2000 Sample**	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Father's Years of Schooling	0.521***	0.621***	0.487***	0.578***
	(0.02200)	(0.02074)	(0.00718)	(0.00781)
Constant	7.067***	5.220***	6.577***	5.544***
	(0.20181)	(0.18799)	(0.08663)	(0.09286)
R2	0.241	0.326	0.284	0.354
Observations	2654	2610	1306800	921106

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Dependent variable: Years of schooling. Years of schooling in Census 2000 was calculated based on the education level completed. Years of schooling in the IFLS was calculated based on highest grade completed in an education level. Data sources: BPS Census 2000, IFLS and IFLS-East.

Table A2: Placebo Test– Exposed Cohorts: 1960 to 1962 and Comparison Cohorts: 1957 to 1959

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	-0.115	-0.182	-0.136	-0.138
	(0.395)	(0.339)	(0.407)	(0.349)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	0.032	-0.086*	0.038	-0.175
	(0.058)	(0.049)	(0.132)	(0.125)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.000	0.010
			(0.012)	(0.011)
R2	0.310	0.387	0.311	0.388
Observations	126511	140868	126511	140868

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.016	0.032	0.011	0.019
	(0.041)	(0.049)	(0.044)	(0.050)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.001	-0.007	0.007	-0.002
	(0.006)	(0.007)	(0.012)	(0.014)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			-0.001	-0.000
			(0.001)	(0.001)
R2	0.146	0.212	0.163	0.234
Observations	126555	140906	126555	140906

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1) Sample corresponds to individuals born between 1957 to 1962. 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. 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. 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: BPS Census 2000 and Duflo (2001).

Table A3: Placebo Test– Exposed Cohorts: 1957 to 1962 and Comparison Cohorts: 1955 to 1957

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1957-1962 × Inpres	0.728	-0.394	0.753	-0.451
	(0.500)	(0.440)	(0.548)	(0.428)
Father's Years of Schooling × Born Between 1957-1962 × Inpres	-0.019	0.037	-0.165	0.032
	(0.080)	(0.063)	(0.196)	(0.160)
Father's Years of Schooling Sq × Born Between 1957-1962 × Inpres			0.018	0.003
			(0.017)	(0.016)
R2	0.312	0.388	0.312	0.389
Observations	135929	154410	135929	154410

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1957-1962 × Inpres	0.106*	0.001	0.100	-0.021
	(0.062)	(0.056)	(0.068)	(0.054)
Father's Years of Schooling × Born Between 1957-1962 × Inpres	-0.013	-0.004	-0.010	0.012
	(0.008)	(0.008)	(0.018)	(0.014)
Father's Years of Schooling Sq × Born Between 1957-1962 × Inpres			0.000	-0.001
			(0.001)	(0.001)
R2	0.148	0.216	0.165	0.238
Observations	135978	154453	135978	154453

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to individuals born between 1955 to 1962. 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. 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. 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: BPS Census 2000 and Duflo (2001).

Table A4: Effects of Inpres Schools on Intergenerational Educational Mobility–Father's Years of Schooling as Family Background (Full Table)

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Father's Years of Schooling	0.450***	0.528***	0.368***	0.482***
	(0.018)	(0.018)	(0.033)	(0.035)
Born Between 1968-1972 × Inpres	1.585***	0.513	1.534***	0.476
	(0.387)	(0.466)	(0.380)	(0.492)
Father's Years of Schooling × Born Between 1968-1972	-0.053***	-0.083***	-0.081***	-0.089***
	(0.013)	(0.017)	(0.021)	(0.024)
Father's Years of Schooling × Inpres	0.257***	0.113*	0.561***	0.311**
	(0.076)	(0.063)	(0.138)	(0.132)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.129**	0.043	-0.255***	-0.020
	(0.054)	(0.062)	(0.086)	(0.099)
Father's Years of Schooling Sq			0.007***	0.004
			(0.002)	(0.003)
Father's Years of Schooling Sq × Born Between 1968-1972			0.001	-0.000
			(0.002)	(0.002)
Father's Years of Schooling Sq × Inpres			-0.027***	-0.018*
			(0.008)	(0.010)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.015**	0.008
			(0.007)	(0.008)
R2	0.322	0.409	0.323	0.409
Observations	1199814	848350	1199814	848350

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Father's Years of Schooling	0.017***	0.028***	0.037***	0.060***
	(0.003)	(0.003)	(0.004)	(0.005)
Born Between 1968-1972 × Inpres	0.204***	0.216***	0.195***	0.197***
	(0.041)	(0.044)	(0.042)	(0.047)
Father's Years of Schooling × Born Between 1968-1972	-0.010***	-0.019***	-0.022***	-0.038***
	(0.001)	(0.002)	(0.003)	(0.002)
Father's Years of Schooling × Inpres	0.043***	0.044***	0.083***	0.076***
	(0.010)	(0.013)	(0.018)	(0.021)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.022***	-0.014**	-0.038***	-0.017*
	(0.006)	(0.006)	(0.011)	(0.010)
Father's Years of Schooling Sq			-0.002***	-0.003***
			(0.000)	(0.000)
Father's Years of Schooling Sq × Born Between 1968-1972			0.001***	0.002***
			(0.000)	(0.000)
Father's Years of Schooling Sq × Inpres			-0.004***	-0.004***
			(0.001)	(0.001)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.002***	0.001*
			(0.001)	(0.001)
R2	0.107	0.172	0.117	0.188
Observations	1200144	848619	1200144	848619

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to individuals 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. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Years of schooling was calculated based on the education level completed. 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: BPS Census 2000 and Duflo (2001).

Table A5: Quadratic CEFs of Exposed and Non-Exposed Cohorts with Father's Years of Education

Panel A: Dependent Variable: Years of Schooling

	Sons		Daughters	
	Not Exposed	**Exposed**	**Not Exposed**	**Exposed**
	(1)	(2)	(3)	(4)
Father's Years of Schooling	0.547***	0.404***	0.641***	0.549***
	(0.02325)	(0.01864)	(0.02678)	(0.02365)
Father's Years of Schooling Sq	0.002	0.005***	-0.000	0.000
	(0.00140)	(0.00115)	(0.00169)	(0.00142)
Constant	5.081***	6.884***	3.682***	5.999***
	(0.09642)	(0.08786)	(0.08994)	(0.09826)
R2	0.277	0.283	0.351	0.346
Observations	128431	1178369	140101	781005

Panel B: Dependent Variable: Primary Completion

	Sons		Daughters	
	Not Exposed	**Exposed**	**Not Exposed**	**Exposed**
	(1)	(2)	(3)	(4)
Father's Years of Schooling	0.058***	0.028***	0.082***	0.040***
	(0.00214)	(0.00133)	(0.00257)	(0.00193)
Father's Years of Schooling Sq	-0.003***	-0.001***	-0.004***	-0.002***
	(0.00011)	(0.00006)	(0.00013)	(0.00009)
Constant	0.666***	0.838***	0.533***	0.772***
	(0.01051)	(0.00703)	(0.01260)	(0.00983)
R2	0.124	0.061	0.194	0.097
Observations	128479	1178707	140143	781270

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Years of schooling was calculated based on the education level completed. Data source: BPS Census 2000.

Table A6: Effects of Inpres Schools on Relative Mobility and Absolute Mobility in Years of Schooling

Panel A: Intergenerational Marginal Association (IGMA)/ Relative Mobility (Sons)				
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
	ΔIGMA	**Normalized ΔIGMA**	**ΔIGMA**	**Normalized ΔIGMA**
$IGMA_0$	-0.2550	-59.35%	-0.0548	-12.76%
$IGMA_6$	-0.0750	-17.46%	-0.0161	-3.75%
$IGMA_9$	0.0150	3.49%	0.0032	0.75%
$IGMA_{12}$	0.1050	24.44%	0.0226	5.25%
$IGMA_{16}$	0.2250	52.37%	0.0484	11.26%
Panel B: Expected Schooling (ES)/ Absolute Mobility (Sons)				
	Highest Intensity(=1)		**Mean Intensity(=0.215)**	
	ΔES	**Normalized ΔES**	**ΔES**	**Normalized ΔES**
ES_0	1.5230	17.40%	0.3274	3.74%
ES_6	0.5330	6.09%	0.1146	1.31%
ES_9	0.4430	5.06%	0.0952	1.09%
ES_{12}	0.6230	7.12%	0.1339	1.53%
ES_{16}	1.2830	14.66%	0.2758	3.15%

Notes: $\Delta IGMA$ represent the effects of the Inpres schools on the change in relative mobility. ΔES represent the effects of the Inpres schools on the change in absolute mobility. 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. $IGMA_k$ and ES_k are calculated at k father's years of education, where $k = 0, 6, 9, 12, 16$. The IGMA and ES values are based on coefficients reported in Panel A of Table A4. The Normalized IGMA/ES is the IGMA/ES value relative to IGMA/ES of children for father's with average years of education (5.57 years), unexposed cohort (born between 1957 to 1962), and zero Inpres intensity for the full sample (combined sample of sons and daughters). Normalized IGMA and ES values are reported in percentage. Data sources: BPS Census 2000 and Duflo (2001).

Table A7: Quadratic CEFs of Exposed and Non-Exposed Cohorts with Mother's Years of Education

Panel A: Dependent Variable: Years of Schooling

	Sons		Daughters	
	Not Exposed	**Exposed**	**Not Exposed**	**Exposed**
	(1)	(2)	(3)	(4)
Mother's Years of Schooling	0.516***	0.397***	0.657***	0.575***
	(0.02806)	(0.02220)	(0.03322)	(0.02682)
Mother's Years of Schooling Sq	0.006***	0.008***	0.002	0.000
	(0.00196)	(0.00153)	(0.00262)	(0.00183)
Constant	5.531***	7.260***	4.152***	6.414***
	(0.10819)	(0.09478)	(0.09426)	(0.10278)
R2	0.240	0.244	0.319	0.312
Observations	105879	1074101	106356	696310

Panel B: Dependent Variable: Primary Completion

	Sons		Daughters	
	Not Exposed	**Exposed**	**Not Exposed**	**Exposed**
	(1)	(2)	(3)	(4)
Mother's Years of Schooling	0.055***	0.027***	0.079***	0.039***
	(0.00216)	(0.00129)	(0.00266)	(0.00192)
Mother's Years of Schooling Sq	-0.003***	-0.001***	-0.004***	-0.002***
	(0.00012)	(0.00006)	(0.00015)	(0.00010)
Constant	0.699***	0.854***	0.581***	0.795***
	(0.00988)	(0.00643)	(0.01190)	(0.00903)
R2	0.105	0.055	0.168	0.087
Observations	105919	1074407	106387	696537

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1). Years of schooling was calculated based on the education level completed. Data source: BPS Census 2000.

Table A8: Effects of Inpres Schools on Intergenerational Educational Mobility– Comparison Cohorts: 1958-1962

Panel A: Dependent Variable: Years of Schooling				
	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	1.597***	0.551	1.534***	0.521
	(0.384)	(0.462)	(0.381)	(0.488)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.123**	0.054	-0.234***	-0.014
	(0.050)	(0.061)	(0.085)	(0.092)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.013**	0.008
			(0.006)	(0.007)
R2	0.321	0.408	0.322	0.408
Observations	1190828	837014	1190828	837014

Panel B: Dependent Variable: Primary Completion				
	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.205***	0.215***	0.196***	0.198***
	(0.041)	(0.043)	(0.042)	(0.046)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.022***	-0.014**	-0.039***	-0.019*
	(0.006)	(0.006)	(0.011)	(0.010)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.002***	0.001**
			(0.001)	(0.001)
R2	0.105	0.170	0.116	0.185
Observations	1191154	837278	1191154	837278

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1) Sample corresponds to individuals 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. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Years of schooling was calculated based on the education level completed. 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: BPS Census 2000 and Duflo (2001).

Table A9: Effects of Inpres Schools on Intergenerational Educational Mobility– Comparison Cohorts: 1959-1962

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	1.666***	0.656	1.570***	0.609
	(0.389)	(0.467)	(0.396)	(0.494)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.131**	0.066	-0.208**	0.017
	(0.051)	(0.061)	(0.097)	(0.091)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.010	0.006
			(0.007)	(0.007)
R2	0.320	0.405	0.321	0.405
Observations	1178986	822791	1178986	822791

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.210***	0.219***	0.199***	0.202***
	(0.041)	(0.041)	(0.043)	(0.044)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.022***	-0.013**	-0.037***	-0.017*
	(0.006)	(0.006)	(0.012)	(0.010)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.002***	0.001*
			(0.001)	(0.001)
R2	0.103	0.166	0.114	0.181
Observations	1179307	823049	1179307	823049

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** p<0.01, ** p<0.05, * p<0.1) Sample corresponds to individuals 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. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Years of schooling was calculated based on the education level completed. 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: BPS Census 2000 and Duflo (2001).

Table A10: Effects of Inpres Schools on Intergenerational Educational Mobility– Comparison Cohorts: 1960-1962

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	1.660***	0.551	1.602***	0.498
	(0.384)	(0.467)	(0.387)	(0.492)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.131**	0.076	-0.247***	0.042
	(0.051)	(0.061)	(0.090)	(0.088)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.014**	0.005
			(0.006)	(0.007)
R2	0.318	0.401	0.319	0.401
Observations	1163653	804747	1163653	804747

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1968-1972 × Inpres	0.207***	0.203***	0.198***	0.189***
	(0.040)	(0.040)	(0.042)	(0.042)
Father's Years of Schooling × Born Between 1968-1972 × Inpres	-0.022***	-0.012**	-0.038***	-0.017*
	(0.005)	(0.006)	(0.011)	(0.009)
Father's Years of Schooling Sq × Born Between 1968-1972 × Inpres			0.002***	0.001**
			(0.001)	(0.001)
R2	0.100	0.159	0.111	0.174
Observations	1163968	805004	1163968	805004

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to individuals 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. Primary completion takes the value of 1 if the child has completed primary and 0, otherwise. Years of schooling was calculated based on the education level completed. 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: BPS Census 2000 and Duflo (2001).

Table A11: Effects of Inpres Schools on Intergenerational Educational Mobility– Partially Exposed Cohorts Only

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1963-1967 × Inpres	0.611**	-0.537*	0.569**	-0.491
	(0.270)	(0.323)	(0.286)	(0.335)
Father's Years of Schooling × Born Between 1963-1967 × Inpres	-0.067	0.100*	-0.112	0.027
	(0.042)	(0.058)	(0.088)	(0.087)
Father's Years of Schooling Sq × Born Between 1963-1967 × 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: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1963-1967 × Inpres	0.094***	0.040	0.082***	0.030
	(0.028)	(0.025)	(0.030)	(0.030)
Father's Years of Schooling × Born Between 1963-1967 × Inpres	-0.013***	-0.003	-0.020**	-0.001
	(0.004)	(0.003)	(0.008)	(0.009)
Father's Years of Schooling Sq × Born Between 1963-1967 × Inpres			0.001**	0.000
			(0.000)	(0.001)
R2	0.128	0.194	0.141	0.214
Observations	433421	400027	433421	400027

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to individuals 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. 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. 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: BPS Census 2000 and Duflo (2001).

Table A12: Effects of Inpres Schools on Intergenerational Educational Mobility–Unrestricted Sample

Panel A: Dependent Variable: Years of Schooling

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1963-1972 × Inpres	1.249***	-0.101	1.197***	-0.155
	(0.346)	(0.376)	(0.338)	(0.388)
Father's Years of Education × Born Between 1963-1972 × Inpres	-0.116**	0.072	-0.244***	0.030
	(0.053)	(0.066)	(0.087)	(0.088)
Father's Years of Education Sq × Born Between 1963-1972 × Inpres			0.015**	0.006
			(0.006)	(0.009)
R2	0.322	0.413	0.322	0.413
Observations	1549978	1169409	1549978	1169409

Panel B: Dependent Variable: Primary Completion

	Linear Model		Quadratic Model	
	Sons	**Daughters**	**Sons**	**Daughters**
	(1)	(2)	(3)	(4)
Born Between 1963-1972 × Inpres	0.166***	0.117***	0.155***	0.094***
	(0.041)	(0.036)	(0.042)	(0.036)
Father's Years of Education × Born Between 1963-1972 × Inpres	-0.020***	-0.010**	-0.034***	-0.008
	(0.006)	(0.005)	(0.011)	(0.008)
Father's Years of Education Sq × Born Between 1963-1972 × Inpres			0.002***	0.001
			(0.001)	(0.001)
R2	0.113	0.182	0.125	0.200
Observations	1550425	1169783	1550425	1169783

Notes: Robust standard errors in parentheses clustered at the district of birth. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$) Sample corresponds to individuals born between 1950 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. 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. 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: BPS Census 2000 and Duflo (2001).