Working Paper in Economics No. 781

Son Preferences and Education Inequalities in India

Heather Congdon Fors and Annika Lindskog

Department of Economics, October 2019, rev. October 2021

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Son Preferences and Education Inequalities in India*

Heather Congdon Fors[1] and Annika Lindskog[2]

School of Economics, Business and Law, University of Gothenburg.

Abstract

We investigate the impact of son preferences in India on gender inequalities in education. We distinguish the impact of preferential treatment of boys from the impact of gender-biased fertility strategies (gender-specific fertility stopping rules and sex-selective abortions). Results show strong impacts of gender-biased fertility strategies on education differences between girls and boys. Preferential treatment of boys has a more limited impact on gender differences. Further, results suggest that gender-biased fertility strategies create gender inequalities in education both because girls and boys end up in systematically different families and because of gender-inequalities in pecuniary investment within families. The extra advantage of the eldest son within the family is small.

JEL codes: D13, I20, J16, O15

Keywords: Son preferences, Gender, Sex-selection, Fertility-stopping rules, Human Capital, Education

[1]ORCID-id 0000-0002-3917-1099

[2] Corresponding author. E-mail annika.lindskog@economics.gu.se; Telephone +46-31-7864412; ORCID-id 0000-0003-1932-8674.

1. Introduction

Son preferences influence a wide range of behaviors in India. This is apparent not least in the country's skewed sex ratios and the large number of missing girls and women (Sen, 1992; Clark, 2000; Klasen & Wink, 2002; Jha et al., 2006; Anderson & Ray, 2010; Jayachandran, 2017; Milazzo, 2018). Studies have also documented unequal health investments between girls/women and boys/men (Arnold et al., 1998; Mishra et al., 2004; Jayachandran & Kuziemko, 2011; Dercon & Singh, 2013; Barcellos et al., 2014), as well as gender gaps in education investment within families (Kingdon, 2005; Azam and Kingdon, 2013; Kaul 2018).

In this paper, we investigate the impact of son preferences in India on gender inequalities in investment and performance in education. Son preferences could translate into education inequalities between the genders due to families who prefer sons investing less in their daughters than in their sons, but also due to the fertility strategies that parents employ to ensure the birth of a son. Indian parents often have strong preferences to have at least one son (with much weaker preferences with regard to the gender of additional children) (Jayachandran, 2017). To ensure the birth of a son, parents may either continue childbearing until they have a son, i.e. gender-specific fertility stopping, or use sex-selective abortions. [1] We will refer to these behaviors broadly as gender-biased fertility strategies. Our main aim is to distinguish the impact of preferential treatment of boys from the impact of gender-biased fertility strategies on education inequalities between girls and boys, which, to the best of our knowledge, has not been done before in a rigorous way.

In families that continue childbearing to have a son, the birth of an additional girl increases family size. This creates gender inequalities in education even if girls and boys are treated equally within the family, due to girls living on average in larger families, which invest less per child, than boys do (Jensen, 2003). Parents' decision to continue childbearing and their decision on how soon they try to have another child will also create inequalities between girls and boys in early-childhood investments (Jayachandran and Pande, 2017; Pörtner, 2020), which has potential consequences for education performance and investment later in life.

[1] Other possible strategies are infanticide and abandonment (Jeffery et al., 1984; Miller, 1987; Sudha and Rajan 1999), but these are likely to be less prevalent.

It has been suggested that sex-selective abortions may reduce gender inequalities, since girls will more often be born into families that actually want them (Goodkind, 1996; Anukriti et al, 2020). However, sex-selective abortions are more common among high caste families and better-educated mothers in well-off families than among economically disadvantaged (Chakraborty and Kim, 2010; Bhalotra and Cochrane, 2010; Jayachandran, 2017; Pörtner, 2020).[2] Therefore, boys are more likely than girls to be born into families with high socioeconomic status (SES), exacerbating gender inequalities.

Our estimation strategy relies on the division of families into those with first-born girls and those with first-born boys. The division of families by the gender of the first-born serves to distinguish families that are more likely to use gender-biased fertility strategies from families that are much less likely to do so. Since sex-selection is not common for first births (Pörtner, 2015; Rosenblum, 2013, 2017), gender of the first-born can be considered random. However, the gender of the first-born has important consequences for the use of gender-biased fertility strategies. Families with a first-born girl might use these strategies to ensure the birth of a son, while families with a first-born son have less reason to use such strategies. In short, we will treat the gender of the first birth as exogenous and we will treat gender of all children as exogenous in the families with first-born boys. We will also use first-born boys' families as the counterfactual without gender-biased fertility strategies when we estimate the impact of these strategies. In the next section, we provide evidence that gender of the first-born is random, and that gender biased fertility strategies are widespread in first-born girl families while they are rare in first-born boy families.

More specifically, to estimate preferential treatment of boys we use the estimated gender effect in families with first-born boys, where gender of later-born children is exogenous. Since the first-born boy himself could be special, we compare later born boys in these families to later born girls. Note that preferential treatment of boys may arise either as a result of parents' favoring sons over

[2] There can be many underlying reasons behind this. One is that they desire smaller families making it less likely to get a son the natural way without exceeding desired fertility. Better-off families could also have stronger reasons to want sons. Edlund (1999) shows that this will be the case if families care about marriage of their children. The skewed sex ratios resulting from sex-selective abortions and excessive mortality among girls makes men the abundant sex on marriage markets. High status men can still expect to get married, while low status men might end up single. For high SES families the choice is thus between a married daughter and a married son, while for low SES families the choice is rather between a married daughter and a potentially unmarried son. This gives low SES families less incentives to abort girls. There could also be differences in availability and affordability of sex-selective abortions, even though it appears to be widely available and cheap (Bhalotra and Cochrane, 2010).

daughters, or due to parents responding to gender differentials in returns to schooling.[3] To estimate gender inequalities that are due to gender-biased fertility strategies, we compare the first-born girls' and the first-born boys' families. We essentially use a difference-in-difference strategy where first-born boys' families are the counterfactual without gender-biased fertility strategies and first-born girls' families are the "treated" group who frequently resort to gender-biased fertility strategies. The interaction term between first-born girl family and gender reveals if girls fare worse in the families that frequently resort to gender-biased fertility than in the families that do not.

Having estimated the impact of gender-biased fertility strategies and preferential treatment of boys for education inequalities, we move on to further investigate sub-channels. First, by adding family fixed effects, we investigate education inequalities between siblings in the same family due to either gender-biased fertility strategies or preferential treatment of boys. We then estimate preferential treatment of the eldest son, who in the earlier literature has been suggested to be particularly advantaged (Jayachandran and Pande, 2017; Kaul, 2018).[4] We do so in a very straightforward way, by adding an eldest son dummy to a within-family estimation that also controls for both gender and birth order.

Last, we investigate whether education inequalities are due to boys more often being born into high SES families and girls into low SES families that also end up having more children. To answer this question we cannot simply estimate between-family regressions since the total between-family inequality will capture education investment responses to child gender in addition to effects of girls and boys ending up in different types of families. Hence, we again compare the first-born girls' and the first-born boys' families, and test whether there are systematic differences in the first-born girls' families regarding where subsequent girls end up in comparison to those where subsequent boys end up. Next, we estimate the consequence of the inferred difference between boys' families and girls' families for educational investment and outcomes. Since there can be additional differences between girls' and boys' families these will be lower bounds.

[3] Further, at least for performance indicators, preferential treatment could be the results of boys receiving more favorable treatment by teachers, for example. Regardless of the root cause, our measure of preferential treatment of boys captures the advantage enjoyed by boys that is not directly related to gender-biased fertility strategies.

[4] The desire of parents to have at least one son might be related to the role of the eldest son in providing for them in old age, taking over family land, and performing important rituals (Mullatti, 1995; Jayachandran and Pande, 2017).

We use a wide range of education indicators, including indicators of performance (completed grades and test scores) as well as of time investment (enrollment and hours spent on school) and pecuniary investment (the private-public school choice and education expenditure). In addition, we estimate the impact on height-for-age, which is a potential link between inequalities in early-life environment and investment and later-life education outcomes. This is interesting since the earlier literature documents implications of gender-biased fertility strategies on early-life investments (Barcellos et al., 2014; Jayachandran and Pande, 2017).

Our results suggest that gender-biased fertility strategies play a greater role than preferential treatment of boys in creating education inequalities between boys and girls, with generally small or insignificant gender differences among later-born children in first-born boys' families. Gender biased fertility strategies create gender gaps among later born children in first-born girl families that are generally larger than the total (average) descriptive inequalities. The resulting gap in completed grades is more than 3 times as large as the descriptive inequality, and the reading gap about twice as large. For enrollment, the resulting gap is about the same size as the descriptive gap though, and for HAZ and math test scores it is lower. Gender gaps that are due to preferential treatment of boys over girls are always smaller than the total (average) descriptive inequalities. The largest, in relative terms, are the gaps in HAZ and math test scores, which are about 2/3 of the descriptive gap, and fairly similar in size to the gap created by gender biased fertility strategies.

Our within-family estimations, on a sample of relatively large families, indicate that among later-born children, girls face a disadvantage in terms of enrollment, total hours spent on schooling, HAZ and math scores compared to boys. In general, preferential treatment of brothers over their own sisters seems to be of at least equal magnitude compared to preferential treatment of boys compared to girls in the main estimations. The only exception is private school attendance, where the within family estimate is small and statistically insignificant. Gender-biased fertility strategies appear to create within-household gender inequalities primarily in pecuniary investments. Later-born girls from first-born girl families are enrolled in private school less often, have lower expenditure on schooling, and somewhat worse writing scores than girls from first-born boy families. These statistically significant effects are about the same size as the estimated impacts of gender-biased fertility strategies in the main estimations.

In comparison to the within-family advantage that boys have over girls, the extra advantage of the eldest son is mostly small and statistically insignificant. However, eldest sons are favored over and above the advantage they have because of their gender and birth order for pecuniary investment.

Further, gender-biased fertility strategies cause systematic differences in the types of families girls live in compared to the types of families boys live in. While we are not able to exactly quantify the implied education inequalities between boys and girls, we can estimate a lower bound. While estimates are statistically significant for most education indicators, these lower bounds are generally smaller than the estimated effects of gender-biased fertility strategies in the main estimations.

We contribute to the literature on son preferences in several ways. This literature has mostly considered effects on early-life outcomes and/or survival (Sen, 1992; Clark, 2000; Klasen & Wink, 2002; Jha et al., 2006; Anderson & Ray, 2010; Milazzo, 2018; Arnold et al., 1998; Mishra et al., 2004; Jayachandran & Kuziemko, 2011; Dercon & Singh, 2013; Barecello et al., 2014; Jayachandran & Pande, 2017). Education investment and outcomes have received less attention. While there are papers documenting gender gaps (Kingdon, 2005; Azam and Kingdon, 2013; Kaul 2018), these typically do not explain the connection to son preferences and gender-biased fertility strategies. The seminal paper by Jensen (2003) shows how gender-specific fertility stopping leads to girls living in larger families than boys, and how this can create inequalities in education investment even if girls and boys were treated equally within families. We extend the analysis of Jensen (2003) by analyzing more recent data, when sex-selective abortions were widely available and fertility rates were much lower. We also use a different strategy to identify gender inequalities that are due to different sources. In particular, gender-biased fertility strategies can also lead to within-household differences, as has been shown by e.g. Jayachandran and Pande (2017) and Barcello (2014). As such, one cannot simply use a within-between family distinction to investigate the impact of gender-biased fertility strategies versus preferential treatment.

As pointed out by Edlund (1999), the more frequent use of sex-selective abortions in high compared to low SES families implies that girls and boys are born into different types of families, thus potentially creating a female under-class. The combination of sex-selective abortions and gender-specific fertility stopping could create large differences in the types of families that girls and boys live in. We investigate this, and the implications thereof on education investment and outcomes, empirically.

We contribute to the literature that estimates gender gaps in education. In this literature, within-family gender gaps are typically interpreted as favoritism (Kingdon, 2005; Kingdon & Azam, 2013; Kaul, 2018). However, gender-biased fertility strategies imply that gender cannot be treated as exogenous in the Indian context. Family fixed effects, which is the strategy typically employed in the literature, is not enough to deal with endogeneity of gender (see e.g. Bharadwaj, 2014). Within family gender gaps could be influenced by gender-biased fertility strategies, and so does not only reflect pure within-family favoritism of the boys.

Several authors suggest strong preferences for at least one son and favoritism of the eldest son rather than sons in general in India (Rosenblum, 2013; Jayachandran, 2017; Jayachandran and Pande, 2017; Kaul, 2018). We directly estimate the advantage of the eldest son while controlling for the birth order of all children in the family.[5]

The remainder of the paper is structured as follows: in section 2 we describe the data and present total gender inequalities and evidence of gender-biased fertility strategies, while in section 3 we present the empirical strategy. In section 4 we present the empirical results, and section 5 concludes the paper.

2. Data and descriptive patterns

2.1 Data and variables

The data comes from two rounds of the India Human Development Survey (IHDS), collected in 2004-05 and 2011-12. This is a nationally representative survey of over 40000 households in India. A particular strength of the data is that it includes an unusually rich set of educational variables. To create a sample of families where we can distinguish between first-born boy and first-born girl families we need the birth order and gender of all children. For this, we use birth histories of women

[5] In spite of the simplicity of this approach, we are to the best of our knowledge the first to use it in a model that fully accounts for birth order. Kaul (2018) uses a within-family model with a control for being the first-born child. To control for the birth order of all children in the family is important because of the systematic difference in the birth order of eldest sons in families where the first-born is a boy (where the eldest son is always the first-born child) versus families where the first-born is a girl (where the eldest son is never the first-born child). It is also important given that the latter family type is more likely to resort to gender-biased fertility strategies.

age 15 to 49, which include all children born to the mother, whether the children are present in the household or not. [6] Our sample includes families with either only children or full siblings.[7]

While all families in our sample can be classified into first-born boy or first-born girl families, we only observe education outcomes of children who live in their maternal household in any of the two surveys. Our main estimation sample is children aged 6 to 17, i.e. children old enough to have started school and young enough to still live in their maternal household. The exceptions are the test scores and height-for-age estimation samples, which, for data collection reasons, only include children age 8-11 in any of the two surveys.[8] If the same child is observed in both survey rounds, we randomly pick one of the observations.[9] We use household weights from the first round to account for the fact that there is some oversampling of certain groups in the data.

While the legal marital age in India is 18, some girls marry and leave the household before their 18th birthday, and we do not have data on their education outcomes. About 3% of children have left the household before age 16, while the percentage of children living elsewhere rises to about 10% for children aged 16 to 17.[10] We therefore run robustness checks on children age 6-15. These are very similar to main results and available on request.

Our main explanatory variable of interest is gender. We create a dummy variable *female* that takes a value of one if the child is a girl, and zero otherwise. We also control for age using age dummies, for survey year, and, in some estimations, absolute birth order, using dummy variables for birth orders one, two, three and four plus.[11]

[6] Multiple birth children (twins, triplets) are excluded from the analysis, since their birth order is not well-defined.
[7] We restrict the sample to families where both the mothers and their husbands have not been previously married. Divorce is very unusual in India, and only 3.9% of children have one or two previously married parents.
[8] The height data is available for all ages, but was collected with a primary focus on children under 5 and between the ages 8 to 11, with other ages included based on availability at the time of the survey. Since there may be systematic differences in which girls and boys are around at the time of the survey we only include the 8-11 year olds in the main analysis. We performed robustness regressions including other age groups (available on request). Results are similar, but sometimes a little less statistically significant.
[9] We also estimated regressions including both observations for children observed in both rounds, which gives essentially identical results (not reported, but available on request).
[10] These figures are based on the second wave of the data, where more information on the age of children living elsewhere is available. The percentage of boys between the ages 6 and 15 living elsewhere is slightly higher than the percentage of girls between the ages 6 and 15 living elsewhere (just over 3% compared to just under 3%). For 16 to 17 year olds, the percentage of boys living elsewhere is about 8.5%, while it is just under 12% for girls.
[11] Birth order 4 plus takes a value of one if the child's birth order is 4 or higher and zero otherwise.

We can broadly categorize our dependent variables into educational performance of the child and educational investment. The indicators of child performance are the test scores on reading, writing and mathematics test administered by interviewers, and the number of completed grades. We use standardized test scores such that they measure age-specific standard deviations from the mean, using the sample population as the age-specific reference.

The indicators of educational investment are *Enrollment*, *Total hours*, *Private school* and *School expenses*. The first two are indicators of time invested in education. *Enrollment* is a dummy variables taking a value of 1 if the child is enrolled in school and zero otherwise. [12] *Total hours* combines all hours related to schooling, including the hours in school, hours of homework and hours of private tuition per week used by the child.[13,14] In our regressions, we set *Total hours* to zero for all children who are not enrolled and estimate on the full sample.

We include two measures of investment into school quality: *Private school* and *School expenses*.[15] *Private school* is a dummy variable taking a value of 1 if the child attends a private school, and 0 if the child attends a public school. *School expenses* measures the cost of school fees, books, uniforms, bus fare and private tuition fees in rupees.[16] In the main analysis, we present estimations conditional on being enrolled, since we believe that these are most straightforward to interpret. We run robustness regressions on the full sample where we code School expenses and Private school as zero for children who are not enrolled in school. Results on these are very similar to the main results and available on request.

In addition to education indicators, we use the height-for-age z-score (HAZ). HAZ is relevant since it is a measure that will capture differences in early life investment and environment (Silventoinen, 2003; Li et al., 2003), and since gender-biased fertility strategies have been found to matter for early life investment. HAZ has been shown to be correlated with both health human capital and

[12] While it is possible that children are enrolled without actually attending school, less than 1% of our enrolled sample report spending zero hours on schooling, and approximately 90% report spending at least 29 hours a week on schooling.
[13] Private tuition depends on pecuniary investment, but we still choose to include it to count all hours equally. If children did not study with private tutors they could instead study alone or with someone else.
[14] We restrict the maximum total hours to 112 per week.
[15] Though the effect of these investments on human capital accumulation remain unclear, parents are likely to make them with the intent to improve the child's human capital.
[16] We restrict maximum school expenditures to 70,000 rupees, which is two times the 99th percentile in the sample.

cognitive and non-cognitive skills (Glewwe et al., 2001; Alderman et al., 2001). The *HAZ* was constructed using the WHO reference tables from 2007.

Descriptive statistics on all variables used in our analysis are presented in Table A1. Enrollment is rather high, at about 85%. The average height-for-age Z score is approximately -1.85. Though this is very low, and quite close to the limit for stunting, it is in line with earlier findings from India (Tarozzi, 2008).

2.2 The total education gender gap

To determine whether preferential treatment of boys in the family and gender-biased fertility strategies are important explanations of education inequalities we need a benchmark for comparison. Table 1 shows the correlations between gender and education outcomes in our sample from regressions that do not distinguish between families with first-born boys and families with first-born girls. Note that these total inequalities are used merely as a benchmark. In particular, we will not decompose the total inequalities into different components. Our aim is to test the importance of specific sources of gender inequalities, focusing on credible identification of these sources. This implies that we will not consider all sources behind the total gender inequalities. For example, we do not estimate the difference in outcomes between first-born girls and first-born boys. While gender of the first-born can be treated as exogenous, we still cannot tell if differences in outcomes are because first-born boys are treated different from first-born girls, or because first-born girl families use gender-biased fertility strategies while first-born boy families do not.

Girls exhibit a disadvantage compared to boys for all outcomes. On average girls are 3.2 percentage points less likely than boys to be enrolled, and they spend 1.8 hours less every week on their schooling. If enrolled, girls are 4.5 percentage points less likely to be in a private school and families spent 415 rupees less on their education. Girls' HAZ is about 0.22 lower than boys' are. The reading test scores are about 0.081 age specific standard deviations lower for girls than for boys, the writing test scores are about 0.067 age specific standard deviations lower. The math test score is about 0.159 age specific standard deviations lower, about twice as large as the disadvantage in reading and writing scores. While girls have completed significantly fewer grades, the effect is so small as to not be particularly meaningful (0.05 years).

Table 1: The total gender inequalities in education - descriptive regressions

	Enrolled	Total hours	Private	Expenses	HAZ
Female	-0.032***	-1.815***	-0.045***	-415.489***	-0.219***
	(0.004)	(0.212)	(0.005)	(45.823)	(0.031)
N	68,841	63,829	58,582	54,027	20,046
Mean	0.851	35.574	0.277	2694.254	-1.850
	Completed grades	Reading	Writing	Math	
Female	-0.051**	-0.081***	-0.067***	-0.159***	
	(0.024)	(0.022)	(0.022)	(0.021)	
N	68,790	19,446	19,282	19,368	
Mean	4.449	-0.070	-0.042	-0.059	

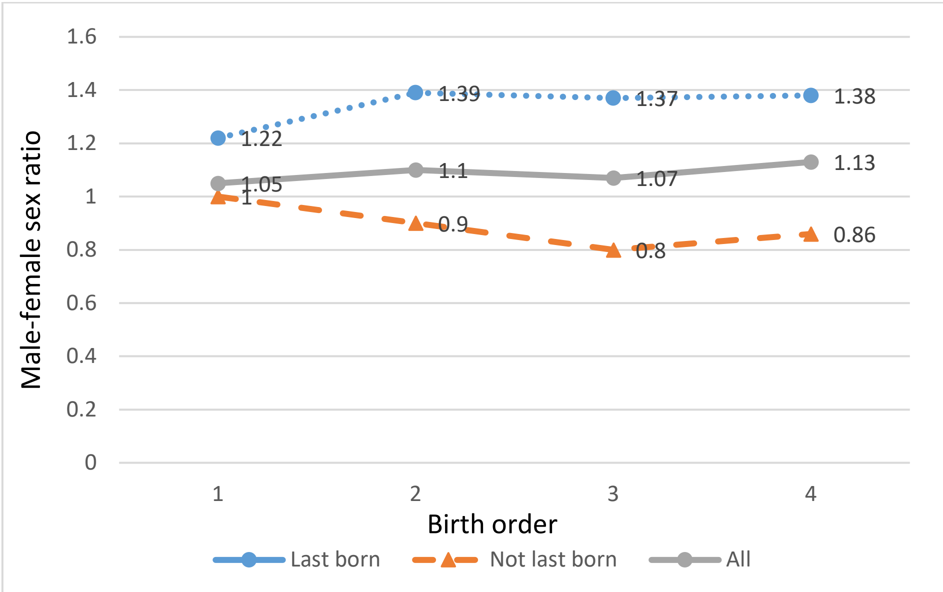
Note: The estimations also include a constant, a full set of child age dummies, and a year dummy. Standard errors, clustered at the sibship level, within parenthesis.

2.2 Evidence on gender-biased fertility strategies in the data

Our empirical strategy assumes that gender of the first-born is random and that gender-biased fertility strategies are used in first-born girl families but not in first-born boy families. This pattern would emerge if families want at least one son, while their preferences regarding the gender of additional children are not strong enough to warrant gender-biased fertility strategies. Earlier studies indeed suggest strong preferences to have at least one son, which leads to increased use of gender-biased fertility strategies when desired fertility is lower (Jayachandran, 2017).

There is clear evidence of gender-biased fertility strategies in the data. Figure 1 shows the number of males relative to females by birth order. When the sample is split between children who are the last-born in the family and those who are not, a striking pattern emerges: the ratio of boys to girls is dramatically larger for last-born children than for children who are not the last-born. This could be either because parents use sex selective abortions before the birth of their last child or because parents continue childbearing if they have a girl. Since we do not observe completed fertility of the mothers these numbers are likely to under-estimate true differences. Some last-born girls might not end up being last born.

Figure 1: Ratio of boys to girls, by birth order



As can be seen in Figure 1 above, the sex ratio for all first-born children is 1.05, which is well within the range that is considered biologically normal (Anderson and Ray, 2010)[17]. Rosenblum (2017) tests for systematic differences in family characteristics that should be exogenous to the gender of the first-born child using the first round of the India Human Development Survey (IHDS), and finds no significant evidence of sex-selection. As we are using both the first and second round of the IHDS, we do similar tests for evidence of sex-selection among first-born children for the data from both rounds. We test for systematic differences in the following family characteristics that should be exogenous to the gender of the first-born child: parental age and education, caste, religion, and whether they live in an urban or rural location. The results are presented in table A2 in the appendix, and show essentially no significant differences between families with first-born girls versus families with first-born boys[18]. Hence, the gender of the first-born can be considered random and there should be no a priori systematic selection into families that have a first-born girl versus a first-born boy.

Our empirical strategy also assumes that gender-biased fertility strategies are used by families with a first-born girl, but not in families with a first-born boy. In table 2 we see that families with a first-

[17] The biologically normal range is 1.03-1.07.

[18] There is a weakly significant difference in belonging to a scheduled caste or scheduled tribe (SC/ST) between first-born girls' and first-born boys' families, with first-born girls' families being on average 1.4% more likely to belong to the category SC/ST.

born girl end up having more children than first-born boy families. This translates into a lower income per capita and a somewhat higher poverty rate. It is clear from the data that gender-biased fertility strategies are used much more in first-born girl than in first-born boy families. The share of girls among (surviving) children of birth order 2 or more is statistically significantly lower in first-born girl families compared to first-born boy families. In first-born boy families, at 0.494 it is just above the natural range at birth (0.483-0.493), while it is much lower in first-born girl families, 0.460. The share of girls among last-born children is dramatically lower in first-born girl than in first-born boy families, 0.376 compared to 0.472. However, at 0.472 it is slightly below the natural level also in first-born boy families. The fact that the total gender ratio is natural in first-born boy families indicate that there are no significant sex-selective abortions, and thus that gender is random. However, the fact that the gender ratio among the last-born is somewhat outside of the natural range indicates that there is still some gender-specific fertility stopping, albeit dramatically less so than in first-born girl families.[19] We discuss our means of addressing this issue in the empirical section below.

Table 2: Differences in non-predetermined outcomes between first-born girl and first-born boy families

	All		First-born girl families		First-born boy families		Difference	
			N	Mean	N	Mean	diff	Std. error
Sibship size	40,662	2.622	19,840	2.744	20,822	2.4505	0.238	0.020
Income pc	40,041	13252	19,527	12,866	20,514	13,618	-752	198
Poor	40,653	0.237	19,836	0.244	20,817	0.231	0.013	0.006
Shares female								
Birth order 1	32,385	0.487						
Birth order 2	29,154	0.476	14,403	0.453	14,751	0.499	-0.046	0.006
Birth order 3	16,759	0.483	9,135	0.472	7,624	0.496	-0.024	0.008
Birth order 4+	15,070	0.469	8,852	0.460	6,218	0.481	-0.021	0.008
Later born birth order 2+	60,983	0.476	32,390	0.460	28,593	0.494	-0.034	0.004
Last born birth order 2+	32,170	0.420	15,901	0.368	16,269	0.470	-0.101	0.005

3. Empirical Strategy

[19] Limited use of gender-biased fertility stopping with no use of sex-selective abortions may be explained by differences in the families employing the two strategies. In particular gender-biased fertility stopping is probably used by families with a weaker preference for small families. In general the decision to abort a girl child or not is fundamentally different from the decision about whether or not to have another child. This is illustrated by the fact that gender-biased fertility stopping appears to occur to some extent also among parents in Western countries with mixed gender preferences (Black et al., 2010), while sex-selective abortions do not appear to be used for this purpose.

To investigate the impact of son preferences on human capital inequalities we identify families with first-born boys and those with first-born girls. The gender of the first-born should be largely exogenous in India despite sex-selective abortions, since these are not common before the birth of the first child (Bhalotra and Cochrane 2010; Jha et al. 2011; Pörtner, 2015; Bharadwaj et al., 2014; Rosenblum, 2013, 2017). In the previous section, we confirmed that this holds in our data; the sex ratio at first births is within the natural range. Earlier literature has used gender of the first-born as a causal estimate of gender in India (Bharadwaj et al., 2014). However, this estimate will capture both preferential treatment of boys and impacts of gender-biased fertility strategies, while our aim is to distinguish these two channels.

As indicated in the previous section, the gender of the first-born leads to important differences between the families in their use of gender-biased fertility strategies. Families with first-born boys have less reason to use either gender-specific fertility stopping rules or sex-selective abortions. This can be exploited in two important ways to learn about the mechanisms behind gender inequalities. First, gender should be as good as random also for later birth orders in the first-born boy families. In the previous section, we showed that the total gender ratios (measured as the share of girls) of additional children in first-born boy families are just above the natural range, suggesting no important role of sex-selective abortions, making gender of each child exogenous. Hence, gender of additional children in first-born boys' families can be considered random in the same way as gender of the first-born. However, the gender coefficient among later-born children in first-born boy families should capture mostly preferential treatment of boys over girls, since these families are much less likely to use gender-biased fertility strategies. Nonetheless, it should be acknowledged that strictly speaking, it measures gender differences that are not due to gender-biased fertility strategies and there could be other reasons than preferential treatment of boys in the family behind these. If families invest less in girls than boys this could for example be because of gender differences in expected returns to education rather than favoring boys because of son preferences (Davies and Zhang, 1995; Kumar, 2013; Rosenblum, 2017). While differences in investment should originate in the family, systematic differences in performance between the genders could also be due to factors outside of the family, such as preferential treatment in schools, or because of differences in expected behaviors in society at large say.

Second, we can compare families with first-born boys to families with first-born girls to find the impact of gender-biased fertility strategies on gender inequalities. In essence, we will use a difference in difference strategy where the gender dummy is interacted with a first-born girl family dummy. The interaction term will capture the additional disadvantage that girls face in families that are likely to employ gender-biased fertility strategies. Hence, the interaction term will be our estimate of the disadvantage that girls face because of fertility strategies.

The main estimation equation is

(1)

$$y_{ist} = \alpha + \beta_1 * female_{is} + \beta_2 * fbg_s + \beta_3 * (fbg_s * female_{is}) + \beta_4 * firstbornboy_{is}$$

$$+ \beta_5 * firstborngirl_{is} + \boldsymbol{age_{ist}}\pi + fbg_s * \boldsymbol{age_{ist}}\pi + \varphi_t + fbg_s * \varphi_t + \varepsilon_{ist}$$

where y_{ist} is outcome y of child i in family s. $female$ is a female dummy, *fbg* is a first-born girl family dummy, $firstbornboy$ is a dummy for first-born boys, $firstborngirl$ a dummy for first-born girls, $\boldsymbol{age}$ are age fixed effects, and φ is a survey dummy. Our main interests is in coefficients β_1 and β_3. β_1 will capture the gender difference among later-born children in first-born boy families, which is our measure of preferential treatment of boys compared to girls (i.e. the disadvantage of girls compared to boys). The interaction term coefficient, β_3, is our measure of impacts of gender-biased fertility strategies. We include a control for being the first-born and a boy or first-born and a girl, but no additional birth order controls. This is because later birth orders are endogenous and related to gender in families that apply gender-biased fertility strategies (Bharadwaj et al., 2014). If birth order matters for education outcomes we still need to control for first births since these are of systematically different genders between the two types of families (even if gender of the first-born is ex ante exogenous).

Descriptive patterns in the previous section indicated marginal gender-specific fertility stopping also in first-born boy families. If so, β_3 would underestimate the true effects of gender-biased fertility strategies. Similarly β_1 may be an upward-biased measure of preferential treatment of boys compared to girls. The sex ratio among last-born children provides a measure of the use of gender-biased fertility strategies (Dalla Zuanna and Leone, 2001; Jayacahndran, 2017). Hence, we use the information in Table 2 of relative skewness of sex ratios among last-born children in first-born boy families compared to in first-born girl families to proxy the bias. Natural sex ratios would imply that girls constituted 0.4878 of children.[20] In first-born girl families, the sex ratio among last-born children is 0.3684 and in first-born boy families it is 0.4701. Hence the relative skewness of sex ratios among last-born children in first-born boy families, in comparison to the difference in skewness between the two types of families, which is what β_3 measure, is (0.4878-0.4701)/(0.4701-0.3684)=0.1740. Hence, our adjusted measure of preferential treatment is $\beta_1 - 0.1740 * \beta_3$ and our adjusted measure of impacts of gender-based fertility strategies is $\beta_3 * 1.1740$.

4. Results

4.1 Main Results

Our main results are presented in Tables 3 for education inputs and Table 4 for education performance. The female coefficient measures the (dis)advantage that girls have compared to boys in first-born boy families, excluding the first-born boy himself. It should be unaffected by sex-selective abortion and can thus be interpreted causally. However, while it should mostly capture preferential treatment of boys compared to girls rather than consequences of gender-biased fertility strategies, there appear to be limited gender-specific fertility stopping also in first-born boy families. We therefore display both the raw coefficient and an adjusted coefficient that should capture only preferential treatment of boys compared to girls. The interaction term between female and first-born girl family measure the additional disadvantage that girls face in families that frequently resort to gender-biased fertility strategies compared to in families that seldom do so. We again present both the raw coefficient, and an adjusted coefficient that should capture the full effect of gender-biased fertility strategies on gender inequalities in education inputs and performance.

Table 3: The effect of preferential treatment of boys versus girls and gender-biased fertility strategies on education inputs

	Enrolled	School hours	Private	Expenses	HAZ
Raw coefficients					
Female	-0.021***	-0.885**	-0.028***	-224.171***	-0.164***

[20] 105 boys per 100 girls, which is in the middle of the natural range 103-107 boys per 100 girls, give this share.

	(0.008)	(0.409)	(0.009)	(68.719)	(0.061)
Female*first-born girl family	-0.024** (0.010)	-1.997*** (0.524)	-0.055*** (0.013)	-461.227*** (99.939)	-0.104 (0.081)
Adjusted coefficients					
Female[a]	-0.017* (0.009)	-0.538 (0.483)	-0.019* (0.011)	-143.917* (81.651)	-0.146** (0.072)
Female*first-born girl family[b]	-0.028** (0.012)	-2.344*** (0.615)	-0.065*** (0.015)	-541.480*** (117.328)	-0.122 (0.095)
R^2	0.11	0.08	0.02	0.08	0.02
N	68,841	63,829	58,582	54,027	20,046
Mean	0.851	35.574	0.277	2694.254	-1.850

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors, clustered at the sibship level, within parenthesis. The regressions also control for first-born girl family, being first-born and a girl, being first-born and a boy, age dummies, a survey dummy and a constant.
[a)] The bias adjusted female coefficient is $\beta_1 - 0.1740 * \beta_3$, where β_1 is the Female coefficient, β_3 is the Female*first-born girl family coefficient, and 0.1740 is the ratio of the skewness in sex-ratio among last borns in first-born boy families to the difference in skewness in the same skewness between first-born boy and first-born girl families.
[b)] The bias-adjusted Female*first-born girl family coefficient is $\beta_3 * 1.1740$.

Gender-biased fertility strategies appear to be more influential in creating inequalities in education inputs between girls and boys than preferential treatment of boys over girls. For raw coefficients, this is the case for hours spent on school, the probability to be in a private school and school expenses. For adjusted coefficients, it is also the case for the probability to be enrolled in school. The only input that deviates from this pattern is HAZ, which is not significantly affected by gender-biased fertility strategies, but is affected by preferential treatment of boys compared to girls. This is noteworthy given the earlier literature that has found impacts of gender-biased fertility strategies on early-life health inputs and HAZ. According to our results, these effects may not persist until ages 8-11 (even if the point estimate is negative). Coefficients for all education inputs indicate that girls are also disadvantaged in families that seldom use gender-biased fertility strategies, however in the case of school hours the adjusted coefficient is statistically insignificant. With the exception of HAZ effects are smaller than from gender-biased fertility strategies.

Table 4: The effect of preferential treatment of boys versus girls and gender-biased fertility strategies on education performance

	Completed grades	Reading	Writing	Math
Raw coefficients				
Female	0.014 (0.044)	-0.035 (0.038)	-0.030 (0.042)	-0.119*** (0.037)

Female*first-born girl family	-0.140**	-0.131**	-0.089	-0.087*
	(0.060)	(0.053)	(0.057)	(0.050)
Bias adjusted coefficients				
Female[a]	0.038	-0.012	-0.014	-0.104**
	(0.052)	(0.045)	(0.050)	(0.044)
Female*first-born girl family[b]	-0.164**	-0.153**	-0.104	-0.102*
	(0.070)	(0.063)	(0.066)	(0.059)
R^2	0.60	0.02	0.01	0.02
N	68,790	19,446	19,282	19,368
Mean	4.449	-0.070	-0.042	-0.059

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors, clustered at the sibship level, within parenthesis. The regressions also control for first-born girl family, being first-born and a girl, being first-born and a boy, age dummies, a survey dummy and a constant.
[a)] The bias adjusted female coefficient is $\beta_1 - 0.1740 * \beta_3$, where β_1 is the Female coefficient, β_3 is the Female*first-born girl family coefficient, and 0.1740 is the ratio of the skewness in sex-ratio among last borns in first-born boy families to the difference in skewness in the same skewness between first-born boy and first-born girl families.
[b)] The bias-adjusted Female*first-born girl family coefficient is $\beta_3 * 1.1740$.

Turning to education performance indicators, gender-biased fertility strategies again appear to be more influential in creating inequalities in education inputs between girls and boys than preferential treatment of boys versus girls, the exception being math test scores.

For completed grades and reading test scores there is a statistically significant disadvantage for girls which is due to gender-biased fertility stopping, perhaps related to the lower investments, but no evidence of preferential treatment. This holds for both the raw and the adjusted coefficients. There are no statistically significant gender effects on writing test scores. For math scores, the raw coefficient indicates a disadvantage for girls which is due primarily to preferential treatment rather than gender-biased fertility strategies, while the adjusted coefficients for the two channels are nearly equal in size. Therefore, the math scores stand out as the only indicator that exhibits a significant role of preferential treatment.[21]

4.2 Investigation of channels - Within family estimations

To investigate whether girls fare worse than their brothers in the same family we next use family fixed effects estimations. Note that preferential treatment of boys is not only a within-family phenomenon, and that consequences of gender-specific fertility stopping is not only a between-family phenomenon. Parents who only have children of one sex could treat these differently than

[21] Whether this is due to preferential treatment by parents or for example by teachers is an open question, and one that we cannot answer with our data.

how they would have treated children of the opposite sex. In addition, the earlier literature suggests that gender-biased fertility strategies could create within-family inequalities, primarily because of less early-life investments when parents try to get pregnant and have a boy soon. Such early-life inequalities could persist into late childhood.

While they are of interest, within-family estimations come with certain caveats. Most importantly, they will out of necessity be identified in a systematically selected sample of rather large families, who have at least one son and one daughter in addition to their first-born child. These families are likely to use less sex-selective abortion than other families, to use gender-biased fertility stopping more than other families, and may in general differ in their son preferences.

In Table 5 and 6 we have added family fixed effects to equation 1. The estimation sample consists of children from families with at least two children in the data, where these children are full-siblings.

Table 5: The effect of preferential treatment of boys versus girls and gender-biased fertility strategies on education inputs – within family estimations

	Enrolled	School hours	Private	Expenses	HAZ
Raw coefficients					
Female	-0.031***	-1.786***	-0.015	-209.305***	-0.244**
	(0.009)	(0.399)	(0.009)	(56.169)	(0.118)
Female*first-born girl family	-0.000	0.407	-0.051***	-390.110***	0.091
	(0.011)	(0.507)	(0.011)	(88.448)	(0.139)
Bias adjusted coefficients					
Female[a]	-0.031***	-1.857***	-0.006	-141.425**	-0.259*
	(0.010)	(0.472)	(0.011)	(67.001)	(0.140)
Female*first-born girl family[b]	-0.001	0.478	-0.060***	-457.990***	0.107
	(0.013)	(0.595)	(0.013)	(103.838)	(0.163)
R^2	0.16	0.13	0.02	0.14	0.07
N	58,373	53,579	47,688	43,383	9,116
Mean	0.849	35.518	0.267	2459.759	1.855

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors, clustered at the sibship level, within parenthesis. The regressions also control for first-born girl family, being first-born and a girl, being first-born and a boy, age dummies, a survey dummy and a constant.

[a)] The bias adjusted female coefficient is $\beta_1 - 0.1740 * \beta_3$, where β_1 is the Female coefficient, β_3 is the Female*first-born girl family coefficient, and 0.1740 is the ratio of the skewness in sex-ratio among last borns in first-born boy families to the difference in skewness in the same skewness between first-born boy and first-born girl families.

[b)] The bias-adjusted Female*first-born girl family coefficient is $\beta_3 * 1.1740$.

Table 6: The effect of preferential treatment of boys versus girls and gender-biased fertility strategies on education performance - within-family estimations

	Completed grades	Reading	Writing	Math
Raw coefficients				
Female	-0.043	-0.100*	-0.069	-0.147***
	(0.043)	(0.057)	(0.065)	(0.052)
Female*first-born girl family	-0.033	-0.066	-0.146*	-0.066
	(0.055)	(0.074)	(0.086)	(0.071)
Bias adjusted coefficients				
Female[a]	-0.038	-0.089	-0.044	-0.135**
	(0.051)	(0.067)	(0.077)	(0.062)
Female*first-born girl family[b]	-0.039	-0.077	-0.171*	-0.077
	(0.065)	(0.087)	(0.101)	(0.083)
R^2	0.61	0.03	0.02	0.03
N	58,324	8,293	8,193	8,262
Mean	4.492	-0.138	-0.108	-0.148

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors, clustered at the sibship level, within parenthesis. The regressions also control for first-born girl family, being first-born and a girl, being first-born and a boy, age dummies, a survey dummy and a constant.
[a)] The bias adjusted female coefficient is $\beta_1 - 0.1740 * \beta_3$, where β_1 is the Female coefficient, β_3 is the Female*first-born girl family coefficient, and 0.1740 is the ratio of the skewness in sex-ratio among last borns in first-born boy families to the difference in skewness in the same skewness between first-born boy and first-born girl families.
[b)] The bias-adjusted Female*first-born girl family coefficient is $\beta_3 * 1.1740$.

Between siblings, girls are disadvantaged with regard to all education inputs except private school enrolment, even in families with little reason to use gender-specific fertility strategies. When we adjust for the possible influence of gender-specific fertility stopping in these families, these differences remain statistically significant, but only weakly so for HAZ. For pecuniary investment, there is a stronger within-family impact of gender-specific fertility stopping. This suggests that families invest less in girls than in their brothers when families become larger and more resource-constrained. Turning to performance indicators in table 6, in spite of higher investments into boys compared to their sisters, they do not generally perform better (at least not statistically significantly so), with the exception of reading (only the raw coefficient) and maths. For writing scores, there is a stronger within-family impact of gender-specific fertility stopping, however these results are only weakly significant.

Several authors suggest strong preferences for at least one son and favoritism of the eldest son rather than sons in general in India (Rosenblum, 2013; Jayachandran, 2017; Jayachandran and Pande, 2017; Kaul, 2018). Hence, next we directly estimate the advantage of the eldest son within the family. We add an eldest son dummy to a within-household estimation that controls for gender (*female*) and birth order (*bo*). Note that the eldest son dummy is not an interaction term since the eldest son could be of any birth order. We also control for a full set of age dummies and survey round.

The model is

$$(2) \quad y_{ist} = \alpha + \beta_1 * eldestson_{is} + \beta_2 * female_{is} + \beta_3 * bo2_{is} + \beta_4 * bo3_{is} + \beta_5 * bo4plus_{is} + \boldsymbol{age_{ist}}\pi + \varphi_t + \gamma_s + \varepsilon_{ist}$$

where $bo2$ is birth order 2, $bo3$ birth order 3, $bo4plus$ birth order 4 or higher, and γ_s the sibship fixed effects. Our coefficient of interest is β_1, which informs us about whether eldest sons are favored over and above any potential benefits they have from being male and of their specific birth order. It should be interpreted as the descriptive advantage of eldest sons within the family. In particular, it should not be causally interpreted since having an eldest son is endogenous and since birth order effects are endogenous to having an oldest son. Results are presented in Tables 7 and 8.

Table 7: The eldest son advantage in education inputs - within-family estimations

	Enrolled	Total hours	Private	Expenses	HAZ
Eldest son	0.008	0.276	0.028***	160.171***	-0.036
	(0.006)	(0.339)	(0.006)	(49.936)	(0.088)
Female	-0.030***	-1.545***	-0.026***	-333.968***	-0.237**
	(0.007)	(0.337)	(0.008)	(47.467)	(0.097)
Second born	-0.024***	-1.728***	-0.009	-234.528***	-0.248***
	(0.006)	(0.293)	(0.008)	(57.070)	(0.081)
Third born	-0.025**	-2.136***	-0.005	-299.368***	-0.463***
	(0.010)	(0.479)	(0.016)	(86.353)	(0.157)
Fourth+ born	-0.008	-2.325***	-0.015	-410.437***	-0.799***
	(0.016)	(0.756)	(0.027)	(122.328)	(0.254)
R^2	0.16	0.13	0.02	0.14	0.08
N	58,373	53,579	47,688	43,383	14,057
Mean	0.849	35.518	0.267	2459.759	1.855

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects. Standard errors, clustered at the sibship level, within parenthesis.* p<0.1; ** p<0.05; *** p<0.01

Table 8: The eldest son advantage in education performance - within-family estimations

	Completed grades	Reading	Writing	Math
Eldest son	0.016	-0.008	0.042	0.043
	(0.035)	(0.045)	(0.051)	(0.044)
Female	-0.051	-0.140***	-0.109**	-0.178***
	(0.036)	(0.045)	(0.052)	(0.042)

Second born	-0.485***	-0.183***	-0.091*	-0.203***
	(0.035)	(0.045)	(0.053)	(0.045)
Third born	-0.776***	-0.330***	-0.116	-0.364***
	(0.060)	(0.074)	(0.090)	(0.073)
Fourth+ born	-0.889***	-0.446***	-0.206*	-0.602***
	(0.099)	(0.111)	(0.123)	(0.104)
R^2	0.61	0.03	0.02	0.04
N	58,324	8,293	8,193	8,262
Mean	4.492	-0.138	-0.108	-0.148

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects.
Standard errors, clustered at the sibship level, within parenthesis.
* p<0.1; ** p<0.05; *** p<0.01

Eldest sons are not favored within families in terms of time investments or HAZ. This indicates that the potential advantage that eldest sons have over their siblings in these measures is explained by their birth order and gender. Eldest sons do, however, appear to be favored in terms of pecuniary measures of investment, over and above what their birth order and gender can explain. Conditional on being enrolled, eldest sons are 2.8 percentage points more likely to attend a private school than what their gender and birth order can explain, and families spend about 160 extra rupees on their education. In the case of education performance, the eldest son coefficient is statistically insignificant and generally small in magnitude. In summary, eldest sons are favored in terms of pecuniary investment, but this does not translate into better education performance as measured here.

4.3 Education gender gaps due to girls and boys being born into systematically different types of families

Do gender-biased fertility strategies imply that girls end up in systematically different types of families that invest less in children's human capital? Since human capital investment is not likely to be fixed, but could respond to child gender we cannot directly test if girls end up in families that invest less. If girls live in families that invest less in education, this could be either because girls ended up in types of families that invest less or because the families invest less when they have more girls. For the same reason, we cannot simply compare models with and without family fixed effects. The total between-family inequality will not only capture the fact that girls and boys end up in different types of families, but also responses in these families to child gender.

We employ a two-step strategy. In the first step, we test whether and how much gender-biased fertility strategies affect the types of families that girls and boys end up in. We do this by estimation of equation 1 on family characteristics, x, that are likely to matter for human capital accumulation. Again, first-born boy families can be seen as providing the counterfactual, not much affected by gender-biased fertility strategies.[22] In the second step, we estimate the correlation between the family characteristics investigated in the first step and the education indicators (Eq. 3).

(3) $$y_{is} = \sum x_{is}\gamma + \varepsilon_{is}$$

Note that it is not important whether the family characteristic's impact on the education outcome is causal or not for our purpose. If girls, for example, more often end up in families where parents have less education, they will on average fare worse than boys, whether the impact of parents education on the education outcome is causal or not. Combining the adjusted interaction term coefficients from estimation of (1) and the coefficients from (3) for each family characteristic x we compute the implied gender inequalities for each outcome y:

(4) $$gap = \sum_{x \in X} \beta_3^{adj}{}_{3,x} * \gamma_x.$$

To get standard errors of the implied gender inequalities, we use cluster bootstrapping. Since we are not likely to include all family characteristics that matter for children's education investment and outcomes, we will estimate lower bounds.

Table 7 below presents results of the first step. We consider mostly predetermined characteristics such as parents' education, religion and caste. Urban residence and total household income are also likely to be largely predetermined. Sibship size is, however, likely to respond to child gender, through gender-specific fertility stopping behavior.

[22] Note that if families with first-born boys do not employ gender-biased fertility strategies and if the dependent variable is predetermined, such that it cannot respond to child gender, the coefficient on the female dummy should not be statistically significant. Appendix Table A3 presents the female coefficients in addition to the interaction terms. The female coefficients are statistically insignificant for all predetermined family characteristics. It is statistically significant for family size, again indicating some use of gender-specific fertility stopping also in first-born boy families.

Table 9: Differences in families that subsequent girls and boys in first-born girl families are born into

	Mother's education	Father's education	Poor	Sibship size	Ideal no children
Raw coefficient	-0.219**	-0.151	0.043***	0.355***	0.076**
	(0.108)	(0.142)	(0.012)	(0.048)	(0.033)
Adjusted coefficient	-0.254**	-0.175	0.050***	0.412***	0.088**
	0.125	0.164	0.014	0.055	0.038
N	72,710	66,636	72,906	73,027	68,915
Mean	3.815	6.122	0.248	3.515	2.628
	Urban	**SC/ST**	**OBC**	**Brahmin**	**Muslim**
Raw coefficient	-0.007	0.004	0.010	-0.004	0.016*
	(0.010)	(0.014)	(0.015)	(0.005)	(0.009)
Adjusted coefficient	-0.009	0.005	0.012	-0.005	0.018*
	0.012	0.016	0.017	0.006	0.011
N	72,197	72,156	72,156	72,156	73,027
Mean	0.325	0.301	0.413	0.049	0.150

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Standard errors, clustered at the sibship level, within parenthesis. The regressions also control for first-born girl family, being first-born and a girl, being first-born and a boy, age dummies, a survey dummy and a constant.

[a)] The bias-adjusted coefficient is the raw coefficient multiplied with1.1740.

Results show that gender-biased fertility strategies result in important differences in the families that boys and girls end up in. In families that are more likely to employ gender biased fertility strategies, i.e. first-born girl families, boys on average end up in families with better-educated mothers than girls. They less often end up in poor families, and they end up in smaller families, where the mother also express a preference for fewer children. Boys are also less often born into Muslim families. There are no statistically significant differences in the likelihood of ending up in families belonging to different caste groups or urban versus rural families.

What are the implied gender inequalities of the fact that girls and boys live in different types of families? Tables A4 and A5 in Appendix show coefficients from regressions of family characteristics on education indicators. These coefficients are then used to predict resulting gender inequalities in education outcomes, displayed in table 10. Note that the gender inequalities in table 10 are lower bounds, since we might miss important family characteristics. Later-born girls in first-born girl families do face a disadvantage on all outcomes. The inequality is about 1/2 of the descriptive total inequality between girls and boys in Table 2 for enrollment and hours spent on schooling, and about 40% for school expenses. For the probability of private schooling and for HAZ it is smaller and not statistically significant (about 15% for private schooling and about 7% for HAZ). The inequality is also about 1/2 of the descriptive total inequality between girls and boys for reading scores and about 2/3 for writing scores, while it is about 2.5 times as large as the total gap for completed grades and about 1/5 of the math score gap. For the math score gap it is not statistically significant.

Table 10: Lower bounds on gender inequalities due to girls and boys ending up in systematically different types of families

	Enrolled	Total hours	Private	Expenses	HAZ
Later-born female in fbg family	-0.014***	-1.0175***	-0.007	-168.091***	-0.016
	(0.003)	(0.223)	(0.006)	(59.358)	(0.017)
	Completed grades	Reading	Writing	Math	
Later-born female in fbg family	-0.125***	-0.043*	-0.040**	-0.036	
	(0.023)	(0.023)	(0.019)	(0.023)	

Gender inequalities are computed according to equation (4) using the adjusted coefficients from Table 9 and the coefficients in Appendix Table A4-A5. Standard errors are computed using cluster bootstrapping.

5. Discussion and Conclusion

We show that son preferences create inequalities in education investment and outcomes between girls and boys and distinguish an impact of preferential treatment of boys compared to girls from an impact of gender-biased fertility strategies. To estimate a gender effect that is due to preferential treatment of boys we use a sub-sample that is unlikely to use gender-biased fertility strategies: families with first-born boys. To identify the impact of gender-biased fertility strategies on education indicators we compare the impact of being a girl in a sub-sample that is likely to use gender-biased fertility strategies (first-born girl families) with a sub-sample that is not likely to do so (first-born boy families).

In essence, we use a difference-in-difference strategy where the female coefficient measure the difference between subsequent girls and subsequent boys born into first-born boy families, i.e. preferential treatment of boys compared to girls. The interaction term between female and first-born girl family measure the additional disadvantage that subsequent girls face when born into a first-born girl compared to a first-born boy family, i.e. impacts of gender-biased fertility strategies. Our data suggest no sex-selective abortion at first births or in first-born boy families, but that they are used in first-born girl families. However, the data suggests limited use of gender-specific fertility stopping also in first-born boy families, even though the practice is much more widespread in first-born girl families. In addition to the raw female and female*first-born girl family coefficients we therefore present estimates adjusted for the bias created by limited use of gender-specific fertility stopping in first-born boy families.

Our results suggest that gender-biased fertility strategies create large education inequalities between girls and boys. Later-born girls age 6-17 born into first-born girl families are 2.8 percentage points less likely to be enrolled than girls in first-born boy families, compared to the mean enrollment of 85%, and they spend 2.3 hours less a week on schooling, compared to the mean of 35.6 hours. They are about 6.5 percentage points less likely to be enrolled in a private school and families spend about 540 rupees less on their education, compared to the means of 27.7% and 2694 rupees, respectively. Moreover, girls perform about 0.15 standard deviations worse on reading tests and have 0.1 standard deviations lower height for age Z scores. Impacts on writing tests and math tests are also negative, but not statistically significant, while the negative impact on completed grades is statistically significant but very small at only 0.16 years, compared to the mean of just under 4.5 years. The effects of gender-biased fertility strategies is particularly large in the case of pecuniary investments, with the effect on private school enrollment approximately one fourth of the sample mean and the effect on school expenditures one fifth of the sample mean.

There are also impacts of preferential treatment of boys compared to girls on gender inequalities, but these are in general smaller than impacts of gender-biased fertility strategies. Our adjusted estimates suggest relatively small and weakly significant gender education gaps in enrollment, private school enrollment, and school expenditures. There are somewhat larger effects on math performance and height for age Z scores, where girls score about 0.12 and 0.16 standard deviations, respectively, less than boys. The raw female coefficient is negative and statistically significant for all education input indicators: enrollment, school hours, private school enrollment, expenses and HAZ, and for math scores.

Gender-biased fertility strategies are primarily used by that half the families with a first-born girl, while preferential treatment of boys compared to girls should affect girls from all family. To judge which source of gender inequality is more influential we can therefore not compare the estimates straight off. However, the gender gaps created by gender-biased fertility strategies are generally more than twice as large as the ones created by preferential treatment, and thus likely to indeed play a greater role in creating gender inequalities in education. The exceptions are gender gaps in HAZ, math test scores, and enrollment, where estimates are of similar size, and consequentially preferential treatment of boys compared to girls may be a more important source of gender inequalities

It is clear from our results that gender-biased fertility strategies, employed by families to ensure the birth of a son, create substantial education inequalities between girls and boys, and that these are more important than education inequalities stemming from preferential treatment of boys compared to girls. Gender-biased fertility strategies could create education inequalities both within families (Jayachandran and Pande, 2017), and between families when girls and boys end up in systematically different types of families (Edlund, 1999). Our investigations of these two mechanisms should only be seen as suggestive, since the within-family estimations are on a selected sample of large families and our between family estimations only provide lower bounds. There appear to be sizeable within-family effects of gender-biased fertility strategies on pecuniary investments, perhaps suggesting that families downgrade educational spending on girls when they become larger and more credit constrained (but again this is identified in a selected sample of large families). There are also sizeable and weakly statistically significant within-family impacts on writing scores. There are no other statistically significant within-family effects of gender-biased fertility strategies. The between-family effects of gender-specific fertility stopping are statistically significant for most outcomes. For education inputs, the estimated effects are statistically significant and larger than within family estimates for time investment (enrollment and hours spent on schooling). The effect on expenses is statistically significant but smaller than the within-family estimate, and there are no statistically significant effects on the probability to attend a private school or on height-for-age Z scores. For performance indicators the estimated effect is statistically significant and larger than the within-family estimate for completed grades. Impacts on test scores are smaller than within-family effects, but still statistically significant for reading and writing scores. In summary, our investigation of whether gender-biased fertility strategies hamper the education of girls mainly through within or between family mechanisms is both suggestive and inconclusive, but both mechanisms appear to matter.

Within-family estimates suggest that girls are treated differently than their brothers with regard to education inputs, in all families, also ones with little reason to employ gender-biased fertility strategies. In general, this does not appear to translate into better performance of boys though. The only performance indictor where boys systematically outperform their sisters is the math test score.

Eldest sons have been suggested to be particularly favored in Indian families. We estimate the advantage that eldest sons enjoy over and above the advantage of their gender and birth order. Eldest sons are favored with regard to private school enrollment and education expenses, but not for other outcomes. They also do not appear to perform better than their siblings do.

It has proven notoriously difficult to create policies to combat gender-biased fertility strategies. While bans on sex-selective abortions might have had some effect (Nandi and Deolalikar, 2013), their prevalence does still appear to have increased over time as families desire smaller families (Anukriti et al., 2020). Sex ratios at birth worsened rather than improved between the 2001 and 2011 censuses (Kulkarni, 2020), and recent data suggest some use of sex-selective abortions even at first-births in recent years (Singh et al, 2021). The central and regional governments in India have also employed various conditional cash transfer schemes to address the issue, but so far there is no evidence that these have had any effect (Sekher and Ram, 2015). Though son preferences so far appear to be sticky in the Indian society there have been some positive developments in other South-east Asian countries such as South Korea (Choi and Hwang, 2020) and Bangladesh (Asadullah et al, 2021), which might give some hope for the future.

If policies are unable to get to the root and reduce the desire to have a son, they can at least do something to even the playing field between girls and boys through education policies that favor girls or socioeconomically disadvantaged families (where girls will more often live compared to boys). India's latest National Education Policy (NEP) 2020 contains many explicit references to gender and gender equality (Ministry of Human Resource Development, 2020). Despite this, concerns have been raised that the policy does not go far enough on this matter (Bhatt, 2020; Sahni, 2020). One concern is that the NEP aims to increase the number of public-private partnerships in the education sector, which may in turn lead to a greater skew towards private schools at the expense of public schools. Our results suggest that girls would be disproportionately negatively affected by such a development.

References

Alderman H, Behrman J, Lavy V, Menon R (2001) Child health and school enrollment: a longitudinal analysis. *Journal of Human Resources* 36:185–205.

Anderson, S., & Ray, D. (2010). Missing women: age and disease. *The Review of Economic Studies*, *77*(4), 1262-1300.

Anukriti, S. (2018) Financial Incentives and the Fertility-Sex Ratio Trade-Off. *American Economic Journal: Applied Economics*, 10(2), 27-57.

Anukriti, S., Bhalotra, S., & Tam, E. H. (2020). On the Quantity and Quality of Girls. Fertility, Parental Investments, and Mortality, World Bank Policy Research Working Paper 9390

Arnold, F., Choe, M. K., & Roy, T. K. (1998). Son preference, the family-building process and child mortality in India. *Population studies*, *52*(3), 301-315.

Asadullah, M. N., Mansoor, N., Randazzo, T., & Wahhaj, Z. (2021). Is son preference disappearing from Bangladesh?. *World Development*, 140, 105353.

Azam, M., & Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development*, *42*, 143-164.

Barcellos, S. H., Carvalho, L. S., & Lleras-Muney, A. (2014). Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?. *American Economic Journal: Applied Economics*, *6*(1), 157-189.

Bhalotra, S. and T. Cochrane (2010). “Where Have All the Young Girls Gone? Identification of Sex Selection in India”, IZA Discussion Paper No. 5381.

Bharadwaj, P., Dahl, G. B., & Sheth, K. (2014). Gender discrimination in the family. *The economics of the family: How the household affects markets and economic growth*, *2*, 237-66.

Bhatt, Neha (2020). “Examining India's new education policy through a gender lens” September 2, 2020. https://www.devex.com/news/examining-india-s-new-education-policy-through-a-gender-lens-98007 Accessed December 2020.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2010). Small family, smart family? Family size and the IQ scores of young men. *Journal of Human Resources*, *45*(1), 33-58.

Chakraborty, T. and S. Kim (2010). "Kinship institutions and sex ratios in India." *Demography* 47(4), 989-1012.

Choi, E. J., & Hwang, J. (2020). Transition of son preference: evidence from South Korea. *Demography*, *57*(2), 627-652.

Clark, S. (2000). "Son preference and sex composition of children: Evidence from India". *Demography*, *37*(1), 95-108.

Congdon Fors, H. and A. Lindskog (2017). "Within-Family Inequalities in Human Capital Accumulation in India: Birth Order and Gender Effects". Department of Economics, School of Business, Economics and Law, University of Gothenburg.

Davies, J. B., & Zhang, J. (1995). Gender bias, investments in children, and bequests. International Economic Review, 795-818.

Dalla Zuanna, G., & Leone, T. (2001). A gender preference measure: the sex-ratio at last birth. Genus, 33-56.

Dercon, S., & A. Singh (2013). "From nutrition to aspirations and self-efficacy: gender bias over time among children in four countries", *World Development*, 45, 31-50.

Desai, Sonalde, Vanneman, Reeve, and National Council of Applied Economic Research, New Delhi (2018). India Human Development Survey (IHDS), 2005. Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR22626.v12

Desai, Sonalde, and Vanneman, Reeve (2018). India Human Development Survey-II (IHDS-II), 2011-12. Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR36151.v6

Edlund, L. (1999). Son preference, sex ratios, and marriage patterns. *Journal of Political Economy*, *107*(6), 1275-1304.

Glewwe P, Jacoby H, King E (2001) Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of Public Economics* 81:345–68.

Goodkind, D. (1996). On substituting sex preference strategies in East Asia: Does prenatal sex selection reduce postnatal discrimination? Population and Development Review, 111-125.

Jayachandran, S. (2017). Fertility decline and missing women. *American Economic Journal: Applied Economics*, 9(1), 118-39.

Jayachandran, S and R Pande (2017) "Why are Indian children so short? The Role of Birth Order and Son Preference", *American Economic Review*, 107(9), 2600-2629.

Jayachandran, S. and I. Kuziemko (2011). “Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India”, *The Quarterly Journal of Economics*, 126(3), 1485–1538

Jensen, R. T. (2003). *Equal Treatment, Unequal Outcomes?: Generating Sex Inequality Through Fertility Behaviour*. Bazaar Chintan.

Jha, P., Kumar, R., Vasa, P., Dhingra, N., Thiruchelvam, D., & R. Moineddin (2006). ”Low male-to-female sex ratio of children born in India: national survey of 1· 1 million households”, *The Lancet*, 367(9506), 211-218.

Kaul, T. (2018). Intra-household allocation of educational expenses: Gender discrimination and investing in the future. *World Development*, *104*, 336-343.

Klasen, S., & Wink, C. (2002). A turning point in gender bias in mortality? An update on the number of missing women. *Population and Development Review*, *28*(2), 285-312.

Kulkarni, P. M. (2020). Sex ratio at birth in India: recent trends and patterns. New Delhi: United Nations Population Fund.

Kumar, A. (2013). “Preference based vs. market based discrimination: Implications for gender differentials in child labor and schooling”. *Journal of Development Economics*, 105, 64-68.

Li H, Stein AD, Barnhart HX, Ramakrishnan U, Martorell R (2003). “Associations between prenatal and postnatal growth and adult body size and composition”. *Am J Clin Nutr*.,77:1498–505.

Milazzo, A. (2018). “Why are adult women missing? Son preference and maternal survival in India”, Journal of Development Economics, 134, 467-484.

Ministry of Human Resource Development, Government of India, 2020. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf Accessed December 2020.

Mishra, V., Roy, T. K., & R.D. Retherford (2004). “Sex differentials in childhood feeding, health care, and nutritional status in India”, *Population and development review*, 30(2), 269-295.

Mullatti, L. (1995). Families in India: Beliefs and realities. *Journal of Comparative family studies*, 11-25.

Nandi, A. and A. B. Deolalikar (2013). ”Does a legal ban on sex-selective abortions improve child sex ratios? Evidence from a policy change in India”, *Journal of Development Economics* 103, 216-228.

Pörtner, C. (2015) “Sex-Selective Abortions, Fertility, and Birth Spacing”, World Bank Policy Research Working Paper No. 7189

Pörtner, C. C. (2020). Birth Spacing and Fertility in the Presence of Son Preference and Sex-Selective Abortions: India's Experience Over Four Decades. *Available at SSRN 3540353*.

Rosenblum, D. (2017) "Estimating the Private Economic Benefits of Sons versus Daughters in India", *Feminist Economics*. 23(1), 77-107.

Rosenblum, D. (2013) "The Effect of Fertility Decisions on Excess Female Mortality in India", *Journal of Population Economics*. 26(1):147-180.

Sahni, Urvashi (2020). "India's National Education Policy 2020: A reformist step forward?" October 2, 2020. https://www.brookings.edu/blog/education-plus-development/2020/10/02/indias-national-education-policy-2020-a-reformist-step-forward/ Accessed December 2020.

Sekher, T. V., & Ram, F. (2015). Conditional cash transfers for girls in India: Assessment of a girl child promotion scheme from beneficiary perspective. International Institute for Population Sciences.

Sen, A. (1992). Missing women. *BMJ: British Medical Journal*, *304*(6827), 587.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of biosocial science*, *35*(02), 263-285.

Singh, A., Kumar, K., Yadav, A. K., James, K. S., McDougal, L., Atmavilas, Y., & Raj, A. (2021). Factors Influencing the Sex Ratio at Birth in India: A New Analysis based on Births Occurring between 2005 and 2016. *Studies in family planning*, 52(1), 41-58.

Tarozzi, A. (2008). Growth reference charts and the nutritional status of Indian children. *Economics & Human Biology*, *6*(3), 455-468.