Online Appendix

1. Dataset Construction: the Florida dataset

1.1. Florida Department of Education Data and Birth Certificates Data

The individual-level administrative data from the Florida Department of Education (FLDOE) Warehouse contains information on K-12 students who attended Florida public schools between the academic year 2002-2003 and 2011-2012. This data was matched with the universe of birth certificates from the Florida Department of Health, comprising all children born in Florida between 1992 and 2002.[1] For the purpose of our analysis, we restrict the sample to the cohorts born from year 1994 onwards, as information on the birth order is not available for children born in 1992 and 1993. The full sample of children enrolled in Florida public schools from 2002-2003 to 2011-2012, and who were born between 1994 and 2002 contains 2,441,705 individuals. Among them, the sample of children whom we are able to match to a birth certificate consists of 1,320,713 individuals.

To obtain a culturally homogenous sample we exclude the children of immigrants by eliminating all students who do not speak English at home and all those students whose mothers were born outside the United States. In the dataset, we are able to reconstruct the language and the maternal country of birth information for 99.92% of the sample.[2] After dropping these individuals, the final sample (which contains students born in Florida, from mothers born in the U.S., and who speak English at home) contains 909,987 individuals.

In order to reconstruct the full fertility history from the birth certificates, we need to be able to observe all children beginning with the firstborn. Therefore we keep only those families where the first child was born in or after 1994. We also eliminate those households where we do not observe all the children between the firstborn and the lastborn.

Starting from the sample of children present in the FLDOE records and born between 1994 and 2002, in order to recreate household composition we match each of them to their mother. From the initial sample of 909,987 individuals, we are able to match 881,798 individuals (96.90%) to their mother ID (In fact note that the sample of children for whom we have information on the mother

[1] The match between the school records and the birth records was implemented by the Florida agencies based on three dimensions: the first and last name, the date of birth, and the social security number. The sample of birth records of children born in Florida from 1994 to 2002 consists of 2,047,633 observations. Of these individuals, 1,652,333 were present in Florida public school data. As reported in Autor et al. (2016), the match rate (81%) is consistent with the percentage of children born in Florida and who attended public school in the State taken from the ACS and the Census over this period. More details on the match are provided in Figlio et al. (2014).

[2] We lose 993 students because the birth country of the mother is coded as unknown (i.e., recorded as "99"), and 7,398 students because their country of birth and/or language spoken at home is coded as unknown (i.e., recorded as "NULL").

country of birth is larger than the one who we are able to match to a maternal ID). We further restrict the sample by dropping those students who belong to households where at least one child speaks at home a language different than English. This leads us to dropping 18,406 children (corresponding to 15,063 households). Furthermore, we drop 159 children, for whom the variable indicating birth order is missing, and 12,666 children (4,812 households) who belong to households where two (or sometimes three) children were recorded as having the same birth order.[3] Finally, we drop 513 children where the data on birth order of the children are inconsistent with the birth year recorded in the school data. The final sample contains 849,295 children (626,628 households).

We use this sample to reconstruct the fertility history of the family to approximate gender biases in the family, following Dahl and Moretti (2008). In order to do this, we face two challenges. First, we need to eliminate the households who have older children who were not enrolled in Florida's public schools between 2002-2003 and 2011-2012: for these children the gender is unknown, as the birth certificate of their younger sibling reports the number of older children but not their gender. This cut further restricts the households and students to respectively 352,138 and 501,274. In order to make sure that fertility is (likely) to be completed, we keep only those households where the probability that the mother has other children after the last one is less than or equal to 10%. The construction of this probability is detailed in section 1.3. of this appendix. Since in the Census there is no indication of whether children of the same age in a given households are twins or only siblings, we further drop from our dataset the families where the mother gave birth to one or more children during the same year (the mother has the same age). This leaves us with 345,968 households and 485,871 children.[4] We also drop mothers who had their first child when they were still teenagers (15 years old or younger). This leaves us with 343,639 households (482,447 children). Finally, we drop the children who come from families with twins because we assume that the arrival of twins might modify future fertility choices. This leads us to dropping 646 observations. Among them, 134,310 households (corresponding to 189,909 children) are the ones likely to have completed fertility, according to our definition.

[3] These are likely due to data entry mistakes as in the case of twins, each child is usually recorded with a unique birth order. In fact, among the sample of twins only 2.6% of them are recorded as having the same birth order (instead, in the rest of the sample 1.1% of children are recorded as having the same birth order)

[4] We also drop 12,300 observations where the birth year of the mother differs across children, or the maternal age at birth was unknown.

1.2. Sample for the test scores regressions

In our sample there are 162,329 students enrolled in grade 6 or higher, corresponding to 630,322 grade/year observations. We keep the student/year observations with non-missing scores in mathematics (we are left with 465,928 observations, corresponding to 153,544 children).[5] If a student repeated a grade, we consider only the year she/he was first enrolled in it (we drop subsequent test scores taken for the same grade). This leads us to drop 4,607 observations (while the number of children remains the same). We also exclude observations corresponding to students who took a math test of grade level different than the one that they are enrolled in. As a result, we drop 973 student/year observations, corresponding to 955 children (among these observations, 98 students are entirely dropped from the dataset; for the rest of the students there are observations corresponding to different years which remain in the dataset). We exclude student/year observations corresponding to students attending a grade two or more years ahead of school (245 student/year observations, which implies dropping 63 children). The mathematics scores are standardized to have mean 0 and standard deviation 1 by test grade level/year across this sample of students. For our baseline regressions, we further drop those households where at least one of the children has an unknown father and drop all the lastborn children in a given family, which leaves us with 139,928 student/year observations (corresponding to 40,177 children). We lose additional 7,102 observations (corresponding to 1,940 children) because we do not have information on median income in zip code of birth; 85 observations (corresponding to 28 children) because we do not have information on the level of education of the mother, 11 observations (4 children) because we do not have information on the marital status of the mother, 111 observations (corresponding to 23 children) because the school id is missing. Our score regression final sample contains 132,619 student/year observations (65,114 girls and 67,505 boys), corresponding to 38,182 unique children (18,512 girls and 19,670 boys). In Table 1, Panel B sample we report sample statistics for the sample of girls. In Appendix Table 4 we report the corresponding statistics for the sample of boys.

[5] In few cases, a student is reported to have more than one score in the mathematics test in a given year. If the repeated scores are identical, we take only one of them, and drop the repetitions. If they are not identical, we assign the student a missing score as we cannot be sure whether these are mistakes, or if they are due to some other reason (for example, the student changed school during the school year, and was administered the test twice). We also assign a missing score whenever the (absolute) difference between the grade attended by the student and the grade level of the test is greater or equal to 2.

2. The probability methodology to estimate completion of fertility

In our FLDOE data, we observe the maternal fertility history only up to 2002 (the last year of the birth certificate data provided in the matched dataset). Thus, we cannot rule out with certainty that the mothers in our sample have additional children born after 2002. To address this issue we use a probabilistic methodology, based on data from the American Community Survey and estimate, for each woman in our sample, the probability that she has indeed completed her fertility by 2002.

The ACS contains information on every child living in the household and their year of birth. For this reason, it has the advantage that the fertility cycle of each mother is more precisely estimated because, differently from the Florida data, for every family observed in the period 2001-2009, information on all previous children born in the family (as long as they live at home) are contained in the dataset. For Florida, for children born in 1994-2002, all the previous children are observed too but the date of birth of the sibling and the gender is known only if they attended Florida public school. Thus, ACS allows us to observe more families with a likely complete fertility.

However, the ACS has two potential problems. First, older children who do not live at home anymore are not accounted for in the data. Second, there is always the possibility that the mother has not completed her fertility. For this reason, we make three assumptions. The first one is that that most women do not have any additional child after 8 or more years from the birth of their last child. This implies that mothers who have children 8 years old, or older, are assumed to have completed fertility. The second one is that we assume that there are no children who have left the households. This is a strong assumption but, if anything, it would imply that we are being too conservative. In fact, by underestimating the number of previous children, we are likely to overestimate the probability that one woman will have more children in the future. The last one is that children leave the household in a sequential way, i.e. older children will leave the household sooner than younger children. We make this assumption to estimate the age of the youngest kid.

In order to construct our probability measures from the ACS, we first eliminate all the families with no children, families where the mother was under 15 at the time of birth of any child (0.41%), and families where the mother was 50 or older at the time of birth of any child (0.04%). Finally, in accordance to our first assumption above, we keep only families for which the youngest observed child is 8 years or older. Note that we identify as "child" those relationships to household head that are "child" and none with any "stepchild", "adopted child", "grandchild", or "foster children".

3. Construction of boy-biased families

Boy-biased families are families where all children are girls except for the last born. For example, for a family with two children, a boy-biased family had a girl as first born and a boy as second born; for a family with three children, a boy-biased family has two girls as first two children and the last child is a boy, and so on. In our analysis, the lastborn is always excluded as there are no lastborn girls in a boy-biased family. As we focus on the sample of girls, by definition every lastborn child would come from non-boy-biased families. In the table below, we list the children combinations for boy-biased families and non-boy-biased ones. We highlight in bold those children who are included in our main regression (Table 3).

	Boy bias=1	Boy bias=0
2 children	**G**B	**G**G
3 children	**GG**B	**G**BB,**GG**G,**G**BG,B**G**B, B**G**G
4 children	**GGG**B	**GGG**G,**GG**BG,**GG**BB,**G**B**G**G, **G**B**G**B,**G**BBG,**G**BBB,B**G**BG, B**G**BB,B**GG**G, B**GG**B

4. Dataset Construction: National Longitudinal Survey of the Youth

We use the 1979 National Longitudinal Survey of Youth (NLSY79) because we are able to observe women's gender role attitudes of mothers in the sample and link them to their children's gender attitudes and performance in math. The original sample includes 12,686 individuals aged between 14 and 22 followed between 1979 and 2014 (yearly interviews until 1994 and biennially after). We focus on the sample of women, which contains 6,283 observations.

Women's gender role attitudes are asked in 1987 and 2004 to all women in the sample.[6] We select three questions:

1) A woman's place is in the home, not in the office or shop;

2) It is much better for everyone concerned if the man is the achiever outside the home and the woman takes care of the home and family;

3) Women are much happier if they stay at home and take care of their children.

For each statement, respondents were asked if they strongly disagreed, disagreed, agreed, or strongly agreed (answers were coded on a scale from 1 to 4). We inverted the answers of the 2004

[6] Some questions were asked also in 1979 and 1982, but we decided to exclude them since at that time the youngest women in the sample were, respectively, 15 and 18 years old. We deem that at this time gender role preferences are not completely formed yet.

wave since the scale was reversed. If the respondent has non-missing values both in 1987 and in 2004, we do a principal component analysis of the answers across years; otherwise, we do a principal component analysis only in the year where no value is missing.[7] Finally, we keep only women who have children, for a total of 4,934 mothers. In our final variable, a higher score means a more gender biased answer.

Starting from 1986, and every two years, a separate survey is administered to the children of the original 1979 NLSY sample (NLSY Children and Young Adults database). Each child is interviewed only for few waves and not every child is interviewed every survey year. Over the years (the last available wave is 2014), 11,521 children were interviewed, corresponding to the 4,934 mothers in our 1979 NLSY sample.

The Young Adults database contains children's attitudes toward women's role and their performance in mathematics measured by the attitudinal test PIAT.[8] We use two sets of variables measuring gender roles attitudes in children: the first set of questions is only asked to children between the age of 10 and 14. The second set of questions is asked to young adults older than 14.

For the first group, we use answers to the six questions asked every year to a subset of interviewees. We pool together answers asked to children aged 10 to 14[9]. Since the same questions are asked in more than one survey, we exploit the longitudinal nature of our dataset and run panel regressions, clustering at the child-level.

We perform a principal component analysis on the answers to the following questions[10]:

1) Girls and boys should be treated the same in school

2) A girl should not let a boy know she is smarter than he is.

3) Competing with boys in school would make a girl unpopular with boys.

4) A girl should pay her own way on dates;

[7] In order to rule out the possibility that an individual has a missing value to an answer which is more important than the others, for every year (i.e. 1987 and 2004) we only keep those women who have non missing value for all of the three questions in at least one year.

[8] We standardize math scores by grade-year, with mean 0 and standard deviation 1 across the full sample.

[9] These questions are also asked also to children aged 14 to 16. We dropped children older than 14, to avoid compositional effects. In addition, the NLSY Children asked these question to children between 14 and 16 only starting from 2002. When we use the full sample, our results do not change. Note how, although the survey specifically says that the children are asked this set of questions in age 10-14, in the data we observe some younger or older children. We cut the ones aged less than 9 (4 observations) and those aged more than 15 (13 observations).

[10] For every year in which such questions are asked, we drop the observations that present at least one missing value to one of the six questions.

5) If there is not enough money for all the children in a family to go to college, the boys should get to go instead of the girls.

6) It is perfectly okay for a girl to ask a boy for a date, even if he has never asked her.

For each statement, the children were asked if they strongly agreed, agreed, disagreed, or strongly disagreed. The possible answer ranged from 1 to 4. For consistency, we recoded all question in such a way that a higher score means a more gender biased answer. We create an index of children gender bias through a principal component analysis of the 6 questions above. In the sample, 6,081 children answer all six questions at least once. Since some children were asked the same questions in multiple years, we have in total 10,103 observations. We lose 155 observations because we do not observe mothers' attitudes, and 1,515 observations because of missing values in the control variables (mother's income, mothers' years of education, mother's birth year, survey year, race, geographical dummy, dummy for being in a relation at the time of the interview, child's age in years).

The final sample counts 8,433 observations (of which 4,126 are boys and 4,307 are girls). This corresponds to 5,380 children (of which 2,668 boys, and 2,712 girls). We present the sample statistics in Table 1, Panel D.

In addition, in the sample there are 7,381 children aged 14 and above (young adults), who answer the same questions on gender attitudes that are asked to their mothers. These young adults are interviewed several times in the survey for a total of 16,761 observations. We lose 333 observations because of missing mothers' attitudes and 2,926 observations because of missing controls. In the end, we have 13,502 observations (6,536 boys and 6,966 girls), corresponding to 6,644 children (3,335 boys and 3,309 girls). Sample statistics for this sub-sample of the 1979 NLSY are shown in Table 1, Panel D.

Finally, we use data on mathematics performance of the female students in NLSY. We keep all the student-year observations for which we have scores in mathematics in grades 6th to 10th (using the same rationale used in the FLDOE data).[11] We start from 10,803 child-year observations and after dropping the ones for which the score in that year was missing we lose 829 observations. We further lose 176 observations because of missing values in the variable women gender role attitudes, and 1,470 because of missing controls (1,466 have missing income, and 4 have missing maternal education). Our final sample contains 8,328 year-grade observations, corresponding to 6,186 students (3,066 boys and 3,120 girls). The sample statistics for this sample are presented in Table 1, Panel C.

[11] Here too, since the dataset is an unbalanced panel, some children appear in the sample multiple times.

References

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2016). "School Quality and the Gender Gap in Educational Achievement." *American Economic Review*, 106(5): 289-295.

Dahl, Gordon B., and Enrico Moretti (2008). "The Demand for Sons." *Review of Economic Studies*, 75(4): 1085–1120.

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth (2014). "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review*, 104(12): 3921-55.

Peri, Giovanni, Kevin Shih, and Chad Sparber (2015). "STEM workers, H-1B visas, and productivity in US cities." *Journal of Labor Economics*, 33(S1): S225-S255.

Variable Description - Main

Name of the variable	*Description*	*Source (and when possible and useful name of the raw variable)*
Firstborn is a girl	A dummy variable equal to 1 if the firstborn child in the household is a girl.	Source: birth certificate, FLDOE Created using raw variables: GENDER_CD
Total number of children	The total number of children had by the child's mother, as reconstructed through the birth certificate.	Source: birth certificate
Two or more children	A dummy variable equal to 1 if the mother had two or more children, equal to 0 otherwise.	Source: birth certificate
Three or more children	A dummy variable equal to 1 if the mother had three or more children, equal to 0 otherwise.	Source: birth certificate
Four or more children	A dummy variable equal to 1 if the mother had four or more children, equal to 0 otherwise.	Source: birth certificate
Family special education	A dummy variable equal to 1 if at least one child was enrolled in the Special education program (excluding the gifted program) in at least one year of our sample.	Source: FLDOE Created using raw variables: PRIMARY_EXCPT_IND
Mother married at first birth	A dummy variable equal to 1 if the mother was married when giving birth to the first child.	Source: birth certificate
Maternal age at first birth	Age of the mother when her first child was born.	Source: birth certificate
Median income in zipcode of birth, USD (family level)	Average income in zipcode of birth across children in the family.	Source: birth certificate, Census
Only child	A dummy variable equal to 1 if the child has no siblings.	Source: birth certificate, Census
Math score	Development scale score in the Mathematics section of the FCAT. The scores are standardized to have mean 0 and standard deviation 1 by test grade level/year across the sample of children enrolled in public school in Florida for whom we are able to reconstruct the fertility history and who took the math test of a level corresponding to the grade they are enrolled in, the first time that they are enrolled in that grade. The scores are standardized by subtracting the mean test score in the sample used for the analysis and by dividing them by the standard deviation for each test grade level-year combination.	Source: FLDOE Created using raw variables: DEV_SCALE_SCORE, SUBTEST_ID, TEST_GRADE_LEVEL, CURRENT_ACADEMIC_YEAR
Boy bias	A dummy equal to 1 if the last born in the family is a boy, and all the older children are girls, 0 otherwise.	Source: birth certificate, FLDOE Created using raw variables: GENDER_CD
Girl Bias		Source: birth certificate, FLDOE

	A dummy equal to 1 if the last born in the family is a girl, and all the older children are boys, 0 otherwise.	Created using raw variables: GENDER_CD
Female	A dummy for whether the student is a boy.	Source: FLDOE Created using raw variables: GENDER_CD
Name of the variable	*Description*	*Source (and when possible and useful name of the raw variable)*
Median income in zip code of birth, (100,000 of $)	The zip code at time of birth (provided by the birth certificates) is matched with zip code income in 1999, obtained from the Census bureau.	Source: birth certificate and Census
Age in months	Assuming the school year starts on September 1st, the variable is calculated as: Academic year*12+8-Student year of birth*12-student month of birth.	Source: FLDOE Created using raw variables: STUDENT_BIRTH_MONTH, STUDENT_BIRTH_YEAR, ENROLLMENT_YEAR
Free or Reduced Priced Lunch	A dummy equal to 1 if the student/year is eligible for free lunch, reduced-price lunch or attends a "provision 2" school and zero otherwise (either the student did not apply or he/she applied but she/he was not eligible).	Source: FLDOE Created using raw variables: LUNCH_STATUS
Mother married at time of birth	A dummy variable equal to 1 if the mother is married at time of giving birth.	Source: birth certificate
Maternal age at birth	Age of the mother when the child was born. The variable was calculated using mother's year and month of birth, and child's year and month of birth.	Source: birth certificate
Special Education	A dummy variable equal to 1 if the variable if the student is enrolled in the special education program and zero otherwise. Gifted students are classified as zero.	Source: FLDOE Created using raw variables: PRIMARY_EXCPT_IND
Mother's educational dummies	We define three dummies for the maternal level of education: high school graduate (years of education is equal to 12), some college (years of education greater than 12 and strictly smaller than 16) and college graduate (years of education greater than or equal to 16).	Source: birth certificate
Family Free Lunch	A dummy variable equal to 1 if at least one child was enrolled in the Free Lunch program in at least one year of our sample, and zero otherwise.	Source: FLDOE Created using raw variables: LUNCH_STATUS

Name of the variable	*Description*	*Source (and when possible and useful name of the raw variable)*
Maternal gender role attitudes	A categorical variable built starting from a set of questions asked to women in years 1987 and 2004. The variable was constructed using the answers to the following question: How much do you agree or disagree with the following statements?" c) A woman's place is in the home, not the office or shop. f) It is much better for everyone concerned if the man is the achiever outside the home and the woman takes care of the home and family. h) Women are much happier if they stay at home and take care of their children. The menu of answers included the following: 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree. We then performed principal component analysis across the three answers of each wave (1987 and 2004). If at least one answer was missing in 1987 (2004), and none were missing in 2004 (1987), the resulting variable is built through principal component analysis on the three questions in 2004 (1987). If none of the answers were missing in 1987 nor in 2004, the final variable was computed though principal component analysis of the answers in 1987 and 2004. If both in 1987 and 2004 there is at least one answer that is missing, a missing value was assigned.	Source: NLSY Created using raw variables: WOMENS_ROLE_000001_1987, WOMENS_ROLE_000006_1987, WOMENS_ROLE_000008_1987, WOMENS_ROLE_000001_2004, WOMENS_ROLE_000006_2004 and WOMENS_ROLE_000008_2004
Gender role attitudes (10 to 14 yrs old)	A categorical variable constructed from a set of questions asked to children aged 10 to 14 years old, in survey waves from 1994 until 2014 (over this period the surveys were administered once every 2 years). It through principal component analysis of the answers to the following questions: How much do you agree or disagree with the following statements? 1) Girls and boys should be treated the same in school; 2) A girl should NOT let a boy know she is smarter than he is; 3) competing with boys in school would make a girl unpopular with boys; 4) A girl should pay her own way on dates; 5) If there is not enough money for all the children in a family to go to college, the boys should get to go instead of the girls. 6) It is perfectly okay for a girl to ask a boy for a date, even if he has never asked her. The menu of answers included the following: 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree. For questions 2, 3 and 5 the scale was reversed. The final value was calculated through principal component analysis across the questions of interests in a given year. A higher value corresponds to higher bias.	Source: NLSY Created using raw variables: CSAS030A, CSAS030B, CSAS030C, CSAS030D, CSAS030E, CSAS030F

Name of the variable	*Description*	*Source (and when possible and useful name of the raw variable)*
Gender role attitudes (over 14 yrs old)	A categorical variable constructed from a set of questions asked to young adults once every 2 years, from 1994 to 2010. It is built through principal component analysis of the answers to the following question: How much do you agree or disagree with the following statements? c) A woman's place is in the home, not the office or shop. f) It is much better for everyone concerned if the man is the achiever outside the home and the woman takes care of the home and family. h) Women are much happier if they stay at home and take care of their children. The menu of answers to this question was the following: 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree. The final value was calculated through principal component analysis of the questions of interests in a given year. A higher value of the variable corresponds to higher bias.	Source: NLSY Created using raw variables: Q16_7C, Q16_7F, Q16_7H
Female		Source: NLSY Created using raw variables: CSEX
Income, USD	Family income.	Source: NLSY Created using raw variables: TNFI_TRUNC
Income (log), USD	log (1+family income)	Source: NLSY Created using raw variables: TNFI_TRUNC
Mother in a relationship	The child's mother is married, has a partner, or is in some other relationship at the time of the survey.	Source: NLSY Created using raw variables: RELSPPTR_YY_XRND (where "YY" stands for the 2-digit code of survey year)
Maternal birth year	Year of birth of the child's mother.	Source: NLSY Created using raw variables: Q1_3_A_Y_1979
Mother's educational dummies	We define three dummies for the maternal level of education: high school graduate (highest grade attended by the mother equal to 12), some college (highest grade attended by the mother greater than 12 and strictly smaller than 16) and college graduate highest grade attended by the mother greater or equal to 16).	Source: NLSY Created using raw variables: HGCREV
Math score	The child's score in the Math PIAT test, standardized by year and grade with mean 0 and standard deviation 1 on the sample of NLSY children. If a child attended the same grade more than once, we only consider the score obtained the first year they are enrolled in a given grade.	Source: NLSY Created using raw variables: MATH
Maternal age at birth	Calculated as the difference between the birth year of the child, and the birth year of the mother.	Source: NLSY CYRB, Created using raw variables: Q1_3_A_Y_1979
Birth order	Child's birth order.	Source: NLSY Created using raw variables: BTHORDR

Age of child (in months)	Age of child in months at the time of the survey.	Source: NLSY Created using raw variables: CSAGE

Appendix - Tables

Table A1: Probability of having additional children

Probability of having more children

Maternal age at birth			Year of birth of lastborn								
1st child	2nd child		2002	2001	2000	1999	1998	1997	1996	1995	1994
20	22		0.451	0.347	0.225	0.142	0.083	0.045	0.021	0.007	0.000
25	27		0.370	0.301	0.198	0.118	0.067	0.035	0.017	0.007	0.000
30	32		0.274	0.219	0.142	0.071	0.036	0.016	0.006	0.002	0.000
35	37		0.157	0.110	0.061	0.027	0.013	0.007	0.003	0.000	0.000
1st child	2nd child	3rd child									
20	22	23	0.331	0.251	0.161	0.074	0.042	0.021	0.010	0.006	0.000
20	22	24	0.305	0.197	0.107	0.047	0.030	0.016	0.004	0.001	0.000
20	22	25	0.250	0.179	0.115	0.063	0.039	0.022	0.004	0.002	0.000
20	22	26	0.192	0.129	0.067	0.051	0.029	0.020	0.013	0.000	0.000
20	22	27	0.185	0.127	0.097	0.054	0.036	0.019	0.013	0.007	0.000
20	22	28	0.171	0.094	0.047	0.017	0.009	0.009	0.009	0.000	0.000
20	22	29	0.097	0.035	0.024	0.009	0.009	0.009	0.000	0.000	0.000
25	27	28	0.306	0.224	0.124	0.051	0.033	0.018	0.009	0.008	0.000
25	27	29	0.259	0.195	0.092	0.052	0.027	0.015	0.005	0.001	0.000
25	27	30	0.240	0.191	0.109	0.066	0.033	0.017	0.002	0.001	0.000
25	27	31	0.155	0.106	0.058	0.038	0.024	0.006	0.001	0.000	0.000
25	27	32	0.126	0.072	0.045	0.022	0.015	0.005	0.003	0.003	0.000
25	27	33	0.091	0.054	0.029	0.015	0.004	0.000	0.000	0.000	0.000
25	27	34	0.065	0.052	0.027	0.007	0.000	0.000	0.000	0.000	0.000
1st child	2nd child	3rd child									
30	32	33	0.285	0.250	0.082	0.055	0.038	0.022	0.006	0.001	0.000
30	32	34	0.227	0.173	0.081	0.035	0.018	0.004	0.001	0.000	0.000
30	32	35	0.139	0.109	0.057	0.018	0.004	0.002	0.001	0.001	0.000
30	32	36	0.093	0.070	0.046	0.018	0.017	0.014	0.000	0.000	0.000
30	32	37	0.066	0.030	0.006	0.000	0.000	0.000	0.000	0.000	0.000
30	32	38	0.050	0.031	0.015	0.007	0.007	0.007	0.007	0.000	0.000
30	32	39	0.049	0.036	0.015	0.000	0.000	0.000	0.000	0.000	0.000
35	37	38	0.218	0.139	0.043	0.036	0.002	0.000	0.000	0.000	0.000
35	37	39	0.160	0.110	0.049	0.030	0.019	0.017	0.017	0.017	0.000
35	37	40	0.099	0.045	0.016	0.006	0.003	0.000	0.000	0.000	0.000
35	37	41	0.039	0.039	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes. The table reports the probability of having an additional child given the number of children already had, the age at which these children we born, and the number of years elapsed since the last birth. In the first Panel, the unit of observation is a mother with a least two children in the American Community Survey, years 2001 to 2009. In the rest of the panels, the unit of observation is a mother with a least three children in the American Community Survey, years 2001 to 2009.

Table A2: Robustness to different thresholds of Pr (Having other children)

	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pr(Other children) ≤ 0.075			Math score					Math score		
Boy bias	-0.027**	-0.013	-0.040**	-0.016	-0.034**	-0.028**	-0.008	-0.043**	-0.013	-0.037**
	(0.013)	(0.020)	(0.017)	(0.022)	(0.017)	(0.013)	(0.022)	(0.018)	(0.023)	(0.017)
Boy bias (standardized beta)	-0.014	-0.007	-0.023	-0.009	-0.019	-0.015	-0.005	-0.025	-0.007	-0.021
Observations	53,780	23,151	30,629	20,772	33,008	49,495	20,800	28,695	18,958	30,537
R-squared	0.323	0.315	0.244	0.330	0.263	0.321	0.317	0.249	0.334	0.264
Pr(Other children) ≤ 0.15										
Boy bias	-0.021**	-0.013	-0.030**	-0.008	-0.031***	-0.023***	-0.010	-0.035***	-0.004	-0.037***
	(0.009)	(0.013)	(0.012)	(0.014)	(0.011)	(0.009)	(0.013)	(0.013)	(0.014)	(0.012)
Boy bias (standardized beta)	-0.012	-0.008	-0.020	-0.005	-0.020	-0.014	-0.006	-0.023	-0.002	-0.024
Observations	84,751	39,488	45,263	35,064	49,687	77,080	35,031	42,049	31,766	45,314
R-squared	0.322	0.296	0.229	0.301	0.253	0.319	0.297	0.232	0.304	0.254
Pr(Other children) ≤ 0.20										
Boy bias	-0.015*	-0.008	-0.025**	-0.002	-0.026**	-0.015*	-0.004	-0.029**	0.003	-0.030***
	(0.008)	(0.011)	(0.011)	(0.012)	(0.010)	(0.008)	(0.012)	(0.012)	(0.013)	(0.011)
Boy bias (standardized beta)	-0.009	-0.005	-0.016	-0.001	-0.017	-0.009	-0.002	-0.019	0.002	-0.019
Observations	98,670	47,016	51,654	41,625	57,045	89,599	41,649	47,950	37,582	52,017
R-squared	0.318	0.285	0.224	0.290	0.247	0.316	0.288	0.227	0.294	0.249

Notes. This table reports OLS estimates, with robust standard errors clustered by student and school. The unit of observation is a student-year. Each Panel runs the same specification shown in Table 3, but applying different thresholds to the probability that the student's mother has had other children we do not observe (the default threshold applied in the analysis is 0.10). In Column (1), the sample includes all girls, excluding lastborns. In Columns (6) to (10), we run the same specifications as in columns (1) to (5), but we restrict the sample to the firstborn child in each family. In Columns (2) and (7), we restrict the sample to families with at least one child enrolled in the Free Lunch program, in at least one year in our sample. In Columns (3) and (8), we restrict the sample to those students who come from families where no child was ever enrolled in the Free Lunch program in any year. In Columns (4) and (9) we restrict the sample to children for whom "Mother high school dropout" or "Mother high school graduate" are equal to 1. In Columns (5) and (10) we restrict the sample to those children with "Mother attended some college" equal to 1, or "Mother graduated from college" equal to 1. The dependent variable measures students' Florida Comprehensive Assessment Test Math score (standardized to have mean 0 and standard deviation 1 by test grade level/year across the sample of children enrolled in public school in Florida for whom we are able to reconstruct the fertility history and who took the math test of a level corresponding to the grade they are enrolled in, the first time that they are enrolled in that grade). "Boy bias" is a dummy variable equal to 1 if the last born in the family is a boy, and all the older children are girls, 0 otherwise. All columns include controls for "Median income in zipcode of birth*100,000 (USD)", "Free Lunch", "Mother married at birth", "Maternal age at birth", "Special Education", "Age (in months)"). Columns (1) to (5) include birth order FE. All columns include year FE, grade FE, school FE, maternal education FE, race FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table A3: Robustness check including lastborn children (FLDOE)

	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Math score					Math score		
Boy bias	-0.021**	-0.017	-0.028**	-0.014	-0.025**	-0.027***	-0.016	-0.037***	-0.014	-0.034**
	(0.010)	(0.015)	(0.013)	(0.016)	(0.013)	(0.010)	(0.016)	(0.014)	(0.017)	(0.013)
Only child	-0.087***	-0.060***	-0.112***	-0.062***	-0.105***	-0.091***	-0.061***	-0.119***	-0.061***	-0.112***
	(0.008)	(0.013)	(0.011)	(0.013)	(0.011)	(0.009)	(0.013)	(0.012)	(0.013)	(0.011)
Median income in zipcode	0.235***	0.229***	0.192***	0.280***	0.200***	0.239***	0.215***	0.209***	0.295***	0.194***
of birth*100,000 (USD)	(0.025)	(0.045)	(0.031)	(0.044)	(0.030)	(0.028)	(0.053)	(0.035)	(0.049)	(0.035)
Free Lunch	-0.151***	-0.081***		-0.121***	-0.181***	-0.147***	-0.077***		-0.118***	-0.180***
	(0.007)	(0.008)		(0.009)	(0.011)	(0.007)	(0.009)		(0.009)	(0.012)
Mother married at birth	0.038***	0.024**	0.056***	0.037***	0.049***	0.042***	0.023**	0.063***	0.041***	0.051***
	(0.008)	(0.010)	(0.014)	(0.010)	(0.013)	(0.009)	(0.011)	(0.015)	(0.011)	(0.014)
Maternal age at birth	0.006***	0.002**	0.008***	0.006***	0.006***	0.005***	0.001	0.007***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Special Education	-0.762***	-0.713***	-0.799***	-0.730***	-0.783***	-0.775***	-0.731***	-0.808***	-0.745***	-0.797***
	(0.014)	(0.019)	(0.022)	(0.019)	(0.021)	(0.016)	(0.021)	(0.025)	(0.021)	(0.026)
Age (in months)	-0.017***	-0.022***	-0.010***	-0.022***	-0.011***	-0.019***	-0.023***	-0.012***	-0.023***	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Birth order FE	YES	YES	YES	YES	YES	-	-	-	-	-
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Maternal Education FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Race FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Boy bias (standardized beta)	-0.010	-0.008	-0.014	-0.006	-0.012	-0.013	-0.007	-0.020	-0.007	-0.018
Observations	182,400	83,319	99,081	81,347	101,053	144,805	66,264	78,541	67,181	77,624
R-squared	0.301	0.272	0.217	0.267	0.243	0.305	0.282	0.227	0.275	0.252

Notes. This table reports OLS estimates, with robust standard errors clustered by student and school. The unit of observation is a student-year. This is the equivalent to Table 3, but it also includes the lastborn child in each family. In Column (1), the sample includes all girls. In Columns (6) to (10), we run the same specifications as in columns (1) to (5), but we restrict the sample to the firstborn in each family. In Columns (2) and (7), we restrict the sample to families with at least one child enrolled in the Free Lunch program, in at least one year in our sample. In Columns (3) and (8), we restrict the sample to those students who come from families where no child was ever enrolled in the Free Lunch program in any year. In Columns (4) and (9) we restrict the sample to children for whom "Mother high school dropout" or "Mother high school graduate" are equal to 1. In Columns (5) and (10) we restrict the sample to those children with "Mother attended some college" equal to 1, or "Mother graduated from college" equal to 1. The dependent variable measures students' Florida Comprehensive Assessment Test Math score (standardized to have mean 0 and standard deviation 1 by test grade level/year across the sample of children enrolled in public school in Florida for whom we are able to reconstruct the fertility history and who took the math test of a level corresponding to the grade they are enrolled in, the first time that they are enrolled in that grade). "Boy bias" is a dummy variable equal to 1 if the last born in the family is a boy, and all the older children are girls, 0 otherwise. Columns (1) to (5) include birth order FE. All columns include year FE, grade FE, school FE, maternal education FE, race FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table A4: Descriptive statistics for the sample of boys, excluding lastborn children (FLDOE)

	(1)	(2)	(3)
	Mean	Std. Dev.	Obs.
Math score	0.444	0.907	67,505
Girl bias	0.461	0.498	67,505
Median income in zipcode of birth*100,000 (USD)	0.491	0.140	67,505
Free Lunch	0.267	0.442	67,505
Mother married at birth	0.841	0.366	67,505
Maternal age at birth	27.274	5.367	67,505
Special Education	0.123	0.329	67,505
Mother high school dropout	0.085	0.279	67,505
Mother graduated high school	0.305	0.461	67,505
Mother attended some college	0.275	0.447	67,505
Mother graduated from college	0.334	0.472	67,505
Age (in months)	157.744	15.823	67,505

Notes. The table reports descriptive statistics for the Florida sample used in Table A5. The unit of observation is a student-year. The sample includes all students born in Florida between 1994 and 2002, in a family where we were able to reconstruct the fertility history without any gap, and for whom we have a score in mathematics. We also exclude students from families where at least one of the children has unknown father. Here, we look only at male students, and we exclude the lastborn child in each family (only children are therefore not included, by definition). "Math score" measures students' Florida Comprehensive Assessment Test math score (standardized to have mean 0 and standard deviation 1 by test grade level/year across the sample of children enrolled in public school in Florida for whom we are able to reconstruct the fertility history and who took the math test of a level corresponding to the grade they are enrolled in, the first time that they are enrolled in that grade). "Girl bias" is a dummy variable equal to 1 if the last born in the family is a girl, and all the older children are boys, 0 otherwise. "Median income in zipcode of birth (USD)" is taken from the 1999 US Census, and it refers to the time of birth of the child. "Free Lunch" is a dummy equal to 1 if the student is enrolled in the Free lunch program in the given academic year. "Mother married at birth" is a dummy variable equal to 1 if the mother was married when the child was born. "Special Education" is a dummy equal to 1 if the student is enrolled in the special education program in the given academic year.

Table A5: Performance in mathematics of boys in families with preferences for girls (FLDOE)

	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college	Only firstborns: All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Math score					Math score		
Girl bias	-0.004	-0.003	-0.009	-0.027	0.009	-0.003	0.004	-0.011	-0.026	0.012
	(0.010)	(0.017)	(0.013)	(0.017)	(0.013)	(0.010)	(0.017)	(0.013)	(0.017)	(0.013)
Median income in zipcode	0.223***	0.241***	0.201***	0.275***	0.196***	0.219***	0.249***	0.197***	0.246***	0.206***
of birth*100,000 (USD)	(0.045)	(0.084)	(0.056)	(0.087)	(0.052)	(0.046)	(0.087)	(0.058)	(0.092)	(0.054)
Free Lunch	-0.136***	-0.064***		-0.113***	-0.156***	-0.131***	-0.062***		-0.106***	-0.159***
	(0.012)	(0.014)		(0.017)	(0.019)	(0.013)	(0.015)		(0.017)	(0.020)
Mother married at birth	0.027	0.020	0.056*	0.023	0.063**	0.031*	0.020	0.064**	0.026	0.059**
	(0.016)	(0.020)	(0.031)	(0.021)	(0.029)	(0.017)	(0.021)	(0.031)	(0.022)	(0.030)
Maternal age at birth	0.005***	0.003	0.005***	0.005**	0.005***	0.004***	0.003	0.005***	0.004**	0.004**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Special Education	-0.674***	-0.678***	-0.660***	-0.660***	-0.679***	-0.683***	-0.684***	-0.669***	-0.665***	-0.689***
	(0.019)	(0.027)	(0.028)	(0.028)	(0.026)	(0.020)	(0.030)	(0.029)	(0.030)	(0.028)
Age (in months)	-0.023***	-0.025***	-0.017***	-0.025***	-0.020***	-0.023***	-0.026***	-0.017***	-0.025***	-0.020***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Birth order FE	YES	YES	YES	YES	YES	-	-	-	-	-
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Maternal Education FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Race FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Girl bias (standardized beta)	-0.002	-0.001	-0.005	-0.015	0.005	-0.001	0.002	-0.007	-0.014	0.007
Observations	67,505	29,325	38,180	26,368	41,137	61,778	26,016	35,762	24,018	37,760
R-squared	0.361	0.364	0.263	0.362	0.299	0.360	0.370	0.269	0.366	0.302

Notes. This table reports OLS estimates, with robust standard errors clustered by student and school. The unit of observation is a student-year. The sample used in Column (1) is the one presented in Appendix Table A4. In Columns (6) to (10), we run the same specifications as in columns (1) to (5), but we restrict the sample to the firstborn child in each family. In Columns (2) and (7), we restrict the sample to families with at least one child enrolled in the Free Lunch program, in at least one year in our sample. In Columns (3) and (8), we restrict the sample to those students who come from families where no child was ever enrolled in the Free Lunch program in any year. In Columns (4) and (9) we restrict the sample to children for whom "Mother high school dropout" or "Mother high school graduate" are equal to 1. In Columns (5) and (10) we restrict the sample to those children with "Mother attended some college" equal to 1, or "Mother graduated from college" equal to 1. The dependent variable measures students' Florida Comprehensive Assessment Test Math score (standardized to have mean 0 and standard deviation 1 by test grade level/year across the sample of children enrolled in public school in Florida for whom we are able to reconstruct the fertility history and who took the math test of a level corresponding to the grade they are enrolled in, the first time that they are enrolled in that grade). "Girl bias" is a dummy variable equal to 1 if the last born in the family is a girl, and all the older children are boys, 0 otherwise. Columns (1) to (5) include birth order FE. All columns include year FE, grade FE, school FE, maternal education FE, race FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table A6: Performance in readings of girls in families with preferences for boys (FLDOE)

						Only firstborns				
	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college	All families	Only families with FRL	Excluding families with FRL	Mother attended HS	Mother attended college
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Reading score					Reading score		
Boy bias	-0.025**	-0.023	-0.030**	-0.023	-0.026*	-0.025**	-0.018	-0.034**	-0.015	-0.028*
	(0.011)	(0.017)	(0.015)	(0.017)	(0.014)	(0.011)	(0.018)	(0.015)	(0.018)	(0.015)
Median income in zipcode	0.219***	0.115	0.214***	0.182**	0.217***	0.226***	0.123	0.215***	0.178*	0.222***
of birth*100,000 (USD)	(0.048)	(0.080)	(0.061)	(0.087)	(0.057)	(0.051)	(0.086)	(0.066)	(0.093)	(0.061)
Free Lunch	-0.168***	-0.092***		-0.143***	-0.187***	-0.165***	-0.087***		-0.139***	-0.186***
	(0.013)	(0.014)		(0.017)	(0.020)	(0.014)	(0.015)		(0.018)	(0.022)
Mother high school grad	0.107***	0.096***	0.108*	0.095***		0.104***	0.089***	0.125**	0.094***	
	(0.021)	(0.023)	(0.061)	(0.023)		(0.023)	(0.026)	(0.063)	(0.024)	
Mother college dropout	0.227***	0.230***	0.196***		-0.236***	0.226***	0.219***	0.220***		-0.240***
	(0.024)	(0.029)	(0.061)		(0.016)	(0.026)	(0.031)	(0.062)		(0.017)
Mother college graduate	0.457***	0.447***	0.413***			0.460***	0.454***	0.439***		
	(0.026)	(0.037)	(0.061)			(0.027)	(0.040)	(0.063)		
Mother married at birth	0.059***	0.044**	0.082**	0.056***	0.068**	0.055***	0.039*	0.079**	0.052**	0.070**
	(0.017)	(0.020)	(0.036)	(0.022)	(0.031)	(0.018)	(0.021)	(0.036)	(0.023)	(0.032)
Maternal age at birth	0.014***	0.009***	0.016***	0.014***	0.013***	0.013***	0.009***	0.015***	0.014***	0.012***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Special Education	-0.749***	-0.713***	-0.768***	-0.740***	-0.738***	-0.758***	-0.713***	-0.789***	-0.741***	-0.756***
	(0.025)	(0.031)	(0.040)	(0.034)	(0.037)	(0.026)	(0.034)	(0.040)	(0.036)	(0.038)
Age (in months)	-0.014***	-0.020***	-0.005***	-0.021***	-0.006***	-0.014***	-0.020***	-0.005**	-0.021***	-0.006***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Birth order FE	YES	YES	YES	YES	YES	-	-	-	-	-
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Grade FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Race FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Boy bias (standardized beta)	-0.014	-0.013	-0.018	-0.013	-0.015	-0.014	-0.011	-0.020	-0.009	-0.016
Observations	65,036	28,944	36,092	25,726	39,310	59,519	25,808	33,711	23,401	36,118
R-squared	0.284	0.293	0.187	0.294	0.211	0.283	0.300	0.191	0.296	0.215

Notes. This table reports OLS estimates, with robust standard errors clustered by student and school. The unit of observation is a student-year. The sample includes all students born in Florida between 1994 and 2002 from a family for whom we were able to reconstruct the fertility history without any gap, and where none of the siblings has unknown father. From these families we keep students enrolled in grades 6th to 10th for whom we have a mathematics score and a reading score. In this table we look only at female students, and we exclude the lastborn child in each family (only children are therefore not included, by definition). Sample statistics for this sample are reported in Table 1, Panel B. In Columns (6) to (10), we run the same specifications as in columns (1) to (5), but we restrict the sample to the firstborn in each family. In Columns (2) and (7), we restrict the sample to families with at least one child enrolled in the Free Lunch program, in at least one year in our sample. In Columns (3) and (8), we restrict the sample to those students who come from families where no child was ever enrolled in the Free Lunch program in any year. In Columns (4) and (9) we restrict the sample to children for whom "Mother high school dropout" or "Mother high school graduate" is equal to 1. In Columns (5) and (10) we restrict the sample to those children with "Mother attended some college" equal to 1, or "Mother college graduate college" equal to 1. The dependent variable measures students' Florida Comprehensive Assessment Test Reading score (standardized to have mean 0 and standard deviation 1 by test grade level/year across the sample of children enrolled in public school in Florida for whom we are able to reconstruct the fertility history and who took the math test of a level corresponding to the grade they are enrolled in, the first time that they are enrolled in that grade). "Boy bias" is a dummy variable equal to 1 if the last born in the family is a boy, and all the older children are girls, 0 otherwise. "Median income in zipcode of birth (USD)" is taken from the 1999 US Census, and it refers to the time of birth of the child. "Free Lunch" is a dummy variable equal to 1 if the student is enrolled in the Free lunch program in the given academic year. "Mother married at birth" is a dummy variable equal to 1 if the mother was married when the child was born. "Special Education" is a dummy equal to 1 if the student is enrolled in the special education program in the given academic year. Columns (1) to (5) include birth order FE. All columns include year FE, grade FE, school FE, race FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table A7: Performance in readings and maternal gender role attitudes (NLSY)

	All	Girls	Boys
	(1)	(2)	(3)
		Reading score (standardized)	
Maternal gender role attitudes	-0.007	-0.017**	0.002
	(0.005)	(0.008)	(0.014)
Female	0.093***		
	(0.014)		
Income (log)	0.031***	0.033**	0.030**
	(0.011)	(0.013)	(0.015)
Mother in a relationship	0.027	0.028	0.023
	(0.025)	(0.032)	(0.031)
Mother high school graduate	0.265***	0.298***	0.233***
	(0.038)	(0.038)	(0.059)
Mother some college	0.437***	0.417***	0.449***
	(0.044)	(0.039)	(0.074)
Mother college graduate	0.649***	0.628***	0.663***
	(0.044)	(0.077)	(0.078)
Maternal age at birth	0.018***	0.023***	0.012**
	(0.003)	(0.005)	(0.005)
Birth order	-0.113***	-0.091***	-0.134***
	(0.007)	(0.013)	(0.018)
Age of child (in months)	-0.010***	-0.010***	-0.010***
	(0.003)	(0.004)	(0.003)
Grade FE	YES	YES	YES
Macro-region FE	YES	YES	YES
Survey year FE	YES	YES	YES
Race FE	YES	YES	YES
Maternal gender role attitudes (standardized beta)	-0.010	-0.027	0.003
Observations	8,201	4,176	4,025
R-squared	0.158	0.151	0.169

Notes. The table reports OLS estimates, with robust standard errors double-clustered at the child and grade level. The unit of observation is a child-year. The sample includes children from NLSY enrolled in grade 6th to 10th, and within the sample, a child may appear in multiple years. In Column (1), the sample includes both girls and boys. Sample statistic for this sample are presented in Table 1, Panel C. In Columns (2) and (3), the sample is restricted respectively to the subset of girls, and to the subset of boys. The dependent variable "Reading score (standardized)" is the child's test score in the reading PIAT test, standardized by survey-year and grade to have mean 0 standard deviation 1 in our sample. The variable "Maternal gender role attitudes" was built based on the answers to the following question, asked to each child's mother in 1987 and 2004: How much do you agree or disagree with the following statements: 1) A woman's place is in the home, not in the office or shop; 2) It is much better for everyone concerned if the man is the achiever outside the home and the woman takes care of the home and family; 3) Women are much happier if they stay at home and take care of their children. The menu of answers to this question was the following: 1: strongly disagree, 2: disagree, 3: agree, 4: strongly agree. A higher value corresponds to a more gender biased family (we recode the answers to the 2004 survey as in that wave the scale was inverted). If at least one answer was missing in 1987 (2004), and none were missing in 2004 (1987), the resulting variable was constructed through the principal component analysis of the three questions in 2004 (1987). If none of the answers were missing in 1987 nor in 2004, the final variable was constructed through the principal component analysis of the answers in 1987 and 2004. If both in 1987 and 2004 there is at least one answer that is missing, the final variable was assigned a missing value. "Female" is a dummy variable (NLSY variable CSEX). "Income, USD" corresponds to net family income (NLSY variable TNFI). "Income (log), USD" was calculated as log(1+Income, USD). "Mother in a relationship" refers to the status at the time of the survey (built from NLSY variable RELSPPTR). Maternal education dummies are built from NLSY variable HGCREV. "Birth order" corresponds to the NLSY variable BTHORDR. "Age of the child (in months)" corresponds to the NLSY variable CSAGE. All regressions include survey year FE, grade FE, macro-region FE, race FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table A8
Choice of a STEM field in college
National Longitudinal Survey of Youth

	All	Girls	Boys
	(1)	(2)	(3)
VARIABLES		Stem field, average	
Maternal gender role attitudes	-0.001	-0.010*	0.011
	(0.005)	(0.006)	(0.009)
Female	-0.142***		
	(0.015)		
Income (log), average	-0.004	0.003	-0.016
	(0.008)	(0.010)	(0.014)
Mother high school graduate	-0.031	0.012	-0.156**
	(0.030)	(0.029)	(0.064)
Mother some college	-0.005	0.028	-0.113*
	(0.032)	(0.033)	(0.065)
Mother college graduate	0.040	0.067	-0.044
	(0.038)	(0.042)	(0.073)
Maternal age at birth	0.001	0.010**	-0.012*
	(0.004)	(0.004)	(0.007)
Mother in a relationship, average	0.020	0.020	0.009
	(0.027)	(0.032)	(0.048)
Birth order	-0.010	-0.021**	0.004
	(0.009)	(0.010)	(0.017)
Maternal gender role attitudes (standardized beta)	-0.004	-0.041	0.037
LHS (mean)	0.254	0.193	0.341
LHS (sd)	0.403	0.358	0.446
Observations	3,043	1,788	1,255
R-squared	0.052	0.060	0.032

Notes. The table reports OLS estimates with robust standard errors. The unit of observation is a child. The sample includes children from NLSY who reported at least one college major in at least one year. In Column (1), the sample includes both girls and boys. In Columns (2) and (3), the sample is restricted respectively to the subset of girls, and to the subset of boys. The dependent variable is built as the average of the dummy variable "stem" across all years for which a college major is reported in the dataset. The dummy variable "stem" is built starting from the NLSY and the classification of stem majors provided by Peri et al. (2015) and available here: http://giovanniperi.ucdavis.edu/uploads/5/6/8/2/56826033/online_appendix.pdf. "Maternal gender role attitudes", the maternal education dummies, and "Birth order" are built as in in Table 5 of the main text. "Income (log), average" and "Mother in a relationship, average" are built as the average across the values of the corresponding variables for a given individual across all available years between 1986 and 2014. The underlying variables are built as described in Table 5 of the main text. All specifications include macro-region FE, initial college year FE, race FE, birth year FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.