I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14007

The Family Origin of the Math Gender Gap Is a White Affluent Phenomenon

Gaia Dossi
David Figlio
Paola Giuliano
Paola Sapienza

DECEMBER 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14007

The Family Origin of the Math Gender Gap Is a White Affluent Phenomenon

Gaia Dossi
London School of Economics

David Figlio
Northwestern University and IZA

Paola Giuliano
UCLA Anderson School of Management and IZA

Paola Sapienza
Northwestern University

DECEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Family Origin of the Math Gender Gap Is a White Affluent Phenomenon*

Previous research has shown that norms around the role of women in society could help explain the gender gap in mathematics and that these norms could be transmitted within the family. Using data from the Florida Department of Education combined with birth certificates we uncover important heterogeneity in the transmission of gender biases within the family. We find that gender role norms can explain the lower performance of girls in mathematics only in relatively affluent White families, whereas they do not apparently matter for the performance of Black girls.

JEL Classification: A13, I20, J16, Z1
Keywords: gender gap in mathematics, cultural transmission

Corresponding author:
Paola Giuliano
Anderson School of Management
University of California, Los Angeles
Los Angeles, CA 90095
USA
E-mail: paola.giuliano@anderson.ucla.edu

* Opinions expressed herein do not reflect the views of the Florida Departments of Education or Health.

The Family Origin of the Math Gender Gap is a White Affluent Phenomenon

By GAIA DOSSI, DAVID FIGLIO, PAOLA GIULIANO AND PAOLA SAPIENZA*

* Dossi: London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom (e-mail: g.g.dossi@lse.ac.uk). Figlio: School of Education and Social Policy, Northwestern University, 2120 Campus Drive, Evanston, IL 60208 (e-mail: figlio@northwestern.edu). Giuliano: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA, 90405 (e-mail: paola.giuliano@anderson.ucla.edu). Sapienza: Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 (e-mail: Paola-Sapienza@kellogg.northwestern.edu). Opinions expressed herein do not reflect the views of the Florida Departments of Education or Health.

Differences in the gender gap in mathematics are strongly correlated with societal gender norms in the U.S. and across the world (Guiso et al., 2008, Nollenberger et al., 2016, Pope and Sydnor, 2010). In this paper, we investigate whether the math gender gap varies by race and socio-economic status. Using a large dataset combining information from the Florida Department of Education and birth records, we show that girls perform worse in mathematics only if they grew up in White families. In Black families, girls systematically outperform boys in mathematics, consistent with Autor et al. (2019). There are at least two non-mutually exclusive explanations for this result. On one hand, it is possible that in disadvantaged families, lower child-rearing inputs (e.g., nutrition, safety in the home, parental attention) have a disproportionate negative effect on the educational and behavioral outcomes of school-age boys relatively to girls. Alternatively, biased gender norms could affect mostly girls in affluent families, because these parents have the educational and financial means to overinvest on the favored gender. This paper investigates the latter hypothesis: whether the greater gender gap in affluent White families is linked to biased educational investments in families with gender biased norms.

Dossi et al. (2019) study the role of gender norms transmitted within the family in explaining the gender gap in mathematics. Using a variety of datasets, the authors find that maternal gender role attitudes are transmitted within the family and could help to partially explain the lower performance of girls in mathematics.

Using a similar methodology, we study whether by measuring the correlation between biased family gender norms and the gender gap in mathematics could help uncover the underlying causes for the differential effect by race. Following Bharadwaj et al. (2015), Dossi et al. (2019) and Dahl and Moretti (2008), we define gender biases within the family using

fertility stopping rules and identify "boy-biased" families as those families where there are only girls and a boy as a last born.

We find that gender role norms can explain the lower performance of girls in mathematics only in relatively affluent White families, whereas they do not apparently matter for the performance of Black girls.

Overall, our results indicate that only boy-biased White families with higher income and maternal education impact negatively girls' math achievement. This result is consistent with the hypothesis that gender biased norms combined with greater educational and financial resources may have the perverse effect of creating a larger gap between boys and girls, whereas families with limited resources are less likely to contribute to the gender gap in mathematics, notwithstanding their gender biases. This result also helps to shed light on Fryer and Levitt's (2010) result that the biggest gender gaps are observed for White girls and for girls in the top quintile of the socioeconomic status.

I. Data and variables of interest

We make use of data from the Florida Education Data Warehouse merged with individual level information coming from the Florida Bureau of Vital Statistics birth certificates. Our cohort is restricted to children born between 1994 and 2002, the period for which we have access to birth certificates. School records contain information on K-12 students who attended Florida public schools between the academic year 2002-2003 and 2011-2012.

Our outcome variables are standardized scores in mathematics (the Florida Comprehensive Assessment Test, or FCAT) from sixth to tenth grade.[1] Controls include children's (age in month, gender, race and whether the child participates in a special education program) and family characteristics (whether the child is eligible to free or reduced lunch or attends a "provision 2" school), all taken from the school records. Education, marital status and maternal age are obtained from the birth certificates together with the zip code at the time of birth.[2] Each regression also

[1] The FCAT (Florida Comprehensive Assessment Test) is the state's high-stakes criterion-referenced test. Students enrolled in public school in grades 3 through 10 are required to take the math portion every year. Students are also tested in reading, but we focus on math because of the broad-based public discussion of women and STEM. The literature has shown that the gender gap in mathematics starts appearing during junior high school (Fryer and Levitt, 2010), for that reason we limit our regressions to students from sixth to tenth grade.

[2] To qualify for free or reduced lunch, the family income has to be respectively below 185% and 130% of the federal income poverty. For details on provision 2 schools see http://www.fns.usda.gov/school-meals/provisions-1-2-and-3. Categories for special education include mentally handicapped, orthopedically, speech, language, or visually impaired, deaf or hard of hearing. It also includes students with emotional or behavioral disabilities, with autistic spectrum disorder and other forms of serious disabilities (such as students with traumatic brain injuries. For maternal education, we define dummies for high school completion, some years of college, and four or more years of college. In the regressions the excluded dummy is high school dropout mothers.

includes birth order, grade, school and year fixed effects. More details on each variable are provided in the Online Appendix.

We use fertility stopping rules to classify families as "boy-biased". We follow Bharadwaj et al. (2015), Dossi et al. (2019) and Dahl and Moretti (2008) who found evidence of parental preferences for boys over girls by showing that the number of children in the US and Florida is significantly higher in families where the first born is a girl. We define "boy-biased" families with a dummy equal to one if all children are girls except for the last born and equal to 0 for all the other families.[3] When comparing girls raised in gender-biased families with those who were not, we always drop the last born from our sample, since, by definition, the last born in a "boy biased" family is always a boy. To rule out the possibility that the results are driven by specific dynamics related to family composition (perhaps the presence of mostly girls in the family does not allow them to learn from boys, who typically do better in mathematics), we present our regressions further restricting the sample to first-born children. As a "placebo" group, we also look at the performance of boys in "girl-biased" families defined with a dummy equal to one if all children are boys except for the last born, who is a girl, and equal to 0 otherwise.

Sample statistics for all our variables of interests are reported in Table A1 of the On-Line Appendix.

II. Estimation results

We start by looking at the gender gap in our sample. Girls perform worse than boys in mathematics in a regression including grade, year, school fixed effects and a large set of individual controls (Table 1, column 1).[4] The beta coefficient is equal to -0.033. This is comparable to three fourths of the difference between students who have a mother who is a high school dropout, and one who is a high school graduate.

As a second step, we split our sample by race. Looking at the heterogeneity by race uncovers interesting differences. The gender gap is pronounced among White girls (the beta coefficient is equal to -0.048, almost double the coefficient of the complete sample). A different picture appears when we investigate the sample of Black girls, who, instead, perform better than boys (although the beta coefficient is relatively small and equal to 0.006). This

[3] For details about the construction of the sample see the On-Line Appendix and Dossi et al. (2019).

[4] Controls include race, age in months, birth order fixed effects, eligibility to free or reduced lunch, a dummy if the student is in a special education program, and several characteristics of the mother (age at birth, marital status and level of education). We also control for the median income of the zipcode at birth.

evidence is consistent with results of Autor et al. (2019): boys fare comparatively worse than girls in education in disadvantaged families. The gender gap is therefore overall driven by girls growing up in White families.

As a second step, we investigate what drives the differences in mathematics performance by race. Dossi et al. (2019) demonstrate that gender biases are important in explaining the differential performance in mathematics of boys and girls. Following their approach we classify families as "boy biased" (or "girl-biased" in our placebo exercise) using fertility stopping rules as explained above.

One interesting question is whether gender biases are relevant only in more affluent families when parents have resources or time to differentiate among their children. Fryer and Levitt (2010) show that girls fall behind boys in math relatively more in families with higher maternal education. We go further and investigate if these differences by socio-economic status are concentrated among those families that have gender biased beliefs about the role of women in societies.

In Table 2, we split the sample of White and Black students using two measures of socio-economic status: eligibility for free or reduced-price lunch and maternal education. Panel A/B present the results for White and Black students, restricting the sample to first-born girls. Column 1 reports the results for the overall sample of White students, where girls perform worse in mathematics (the beta coefficient is -0.022). When we further split the sample by eligibility to free lunch, the results appear to be driven by relatively affluent families (the coefficient is significant when we restrict the sample to families not eligible to free and reduced lunch with a beta coefficient of -0.031). The differences are similar when we split the sample by education of the mother (the beta coefficient is -0.030 for girls whose mothers have at least some college education).

These differences are not present among girls belonging to Black families, where the effect of gender bias is null and there is no significant difference across socio-economic status.

Table 3 reports our placebo exercise where we look at performance of boys, by race and socio-economic status. Gender biases inside the families are not correlated with the performance of boys, reducing the possibility that our measure of gender bias based on fertility stopping rule is picking up some other omitted variables related to specific family dynamics across siblings of the same gender.

Overall our results indicate that gender biases inside the family matter for White affluent families. In Black families girls do perform better than boys but the results cannot be explained by differences in gender roles

transmitted by their parents. Thus, it is likely that the relative underperformance of boys in Black families is due to the fact that lower child-rearing inputs or other aspects of disadvantage have a disproportionate negative effect on boys in these families (Autor et al., 2019).

III. Conclusions

Previous research has shown that norms around the role of women in society could help explain the gender gap in mathematics, and that these norms could be transmitted within the family.

Using data from the Florida Department of Education combined with birth certificates we uncover important heterogeneity in the transmission of gender biases within the family. Gender-biased attitudes (proxied in this paper by fertility stopping rules) are correlated with lower performance in mathematics only in White affluent families. Girls do better in Black families, where the results do not appear to be driven by a different transmission of gender norms.

REFERENCES

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2019. “Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes.” *American Economic Journal: Applied Economics*, 11(3), 338-381.

Bharadwaj, Prashant, Gordon B. Dahl, and Ketki Sheth (2015). “Gender Discrimination in the Family,” “The Economics of the Family”, ABC-Clio Publishers, pp. 237-266.

Block, J.H., 1976. "Issues, Problems, and Pitfalls in Assessing Sex Differences: A Critical Review of The Psychology of Sex Differences", Merrill-Palmer Quarterly of Behavior and Development, 283-308.

Carlana, M. (2019), “Implicit Stereotypes: Evidence from Teachers’ Gender Bias”, *Quarterly Journal of Economics*, Volume 134(3), pp. 1163–1224.

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

Dossi, Gaia, David Figlio, Paola Giuliano and Paola Sapienza. 2019. “Born in the Family: Preferences for Boys and the Gender Gap in Math”, NBER WP 25535.

Fryer, Roland G., and Steven D. Levitt (2010). "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics*, 2010(2.2): 210-40.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales (2008). “Culture, Gender and Math.” *Science*, 320(5880): 1164–65.

Hoffman, L. W., 1977. "Changes in Family Roles, Socialization, and Sex Differences", American Psychologist, 32(8), 644-657.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. Psychological Bulletin, 109, 267–296. doi:10.1037/0033-2909.109.2.267

Nollenberger, N., N. Rodriguez-Planas and A. Sevilla (2016). "The Math Gender Gap: The Role of Culture", *American Economic Review*, 106 (5): 257-261.

Pope, Devin G., and Justin R. Sydnor (2010). "Geographic Variation in the Gender Differences in Test Scores", *Journal of Economic Perspectives*, 24(2): 95-108.

TABLE 1— GENDER GAP IN MATHEMATICS BY RACE, FLORIDA DEPARTMENT OF EDUCATION

	All	White	Black
	(1)	(2)	(3)
		Math score	
Female	**-0.061*** **	**-0.083*** **	**0.011***
	(0.003)	**(0.004)**	**(0.006)**
Female (standardized beta)	-0.033	-0.048	0.006
Observations	703,654	489,903	154,253
R-squared	0.355	0.291	0.276

Notes. This table reports OLS estimates, with robust standard errors clustered at the student and school level. The unit of observation is a student-year. The sample includes all students born in Florida between 1994 and 2002, enrolled in grades 6th to 10th in a Florida Public School, and for whom we have a mathematics test score. Sample statistics for this sample are reported in Appendix Table A1. In Column (2), we restrict the sample to White students. In Column (3) we restrict the sample to Black students. The dependent variable measures students' Florida Comprehensive Assessment Test (FCAT) math score in a given grade (standardized with mean 0 and variance 1 over the population for a given grade and year). In all columns, we control for median income in zipcode of birth in USD (taken from the 1999 US Census); a "Free Lunch" dummy variable equal to 1 if the student is enrolled in the Free lunch program in the given academic year; a dummy for "Mother married at birth" equal to 1 if the mother was married when the child was born; a dummy for "Special Education" equal to 1 if the student is enrolled in the special education program in the given academic year. Column (1) includes race FE. All columns include year FE, grade FE, school FE. ***, **, and * indicate significance respectively at the 1%, 5%, and 10% levels.

TABLE 2— GIRLS' PERFORMANCE IN MATHEMATICS AND GENDER BIASES, HETEROGENEITY BY RACE AND SOCIO-ECONOMIC STATUS, FLORIDA DEPARTMENT OF EDUCATION

	All families	Families with FRL	Families without FRL	Mother attended HS	Mother attended college
Panel A: White, only firstborn girls					
	(1)	(2)	(3)	(4)	(5)
			Math score		
Boy bias	**-0.035*** **	**-0.013**	**-0.047*** **	**-0.019**	**-0.045*** **
	(0.011)	**(0.019)**	**(0.015)**	**(0.019)**	**(0.015)**
Boy bias (standardized beta)	-0.022	-0.008	-0.031	-0.012	-0.030
Observations	50,402	19,223	31,179	19,062	31,340
R-squared	0.297	0.310	0.232	0.313	0.229
Panel B: Black, only firstborn girls					
Boy bias	**0.009**	**0.002**	**0.055**	**-0.031**	**0.015**
	(0.041)	**(0.047)**	**(0.128)**	**(0.070)**	**(0.063)**
Boy bias (standardized beta)	0.005	0.001	0.036	-0.018	0.010
Observations	5,455	4,426	1,029	2,801	2,654
R-squared	0.464	0.473	0.725	0.528	0.551

Notes. This table reports OLS estimates, with robust standard errors clustered at the student and school level. 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 none of the children has unknown father. From these families we keep students enrolled in grades 6th to 10th for whom we have a math test score. In this table we look only at firstborn female students, excluding only children. Sample statistics for this sample are reported in Appendix Table A2, Panel A. In Column (2), the sample is restricted to families with at least one child eligible for free or reduced-price lunch (FRL), in at least one year in our sample. In Column (3), the sample is restricted to students from families where no child ever was eligible for free or reduced-price lunch in any year. In Column (4) the sample is restricted to students whose mother is either a high school dropout or a high school graduate (never enrolled in college). In Column (5), the sample is restricted to students whose mother attended college. The dependent variable measures the student score in the Florida Comprehensive Assessment Test (FCAT) in mathematics in a given grade (the score is standardized with mean 0 and variance 1 over the population for a given grade and year). "Boy bias" is a dummy equal to 1 if the last born child in the family is a boy, and all the older children are girls, equal to 0 otherwise. The set of controls is identical to the one described in Table 1. All columns include year FE, grade FE, school FE. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

TABLE 3— BOYS' PERFORMANCE IN MATHEMATICS AND GENDER BIASES, HETEROGENEITY BY RACE AND SOCIO-ECONOMIC STATUS, FLORIDA DEPARTMENT OF EDUCATION

	All families	Families with FRL	Families without FRL	Mother attended HS	Mother attended college
Panel A: White, only firstborn boys					
	(1)	(2)	(3)	(4)	(5)
			Math score		
Girl bias	**-0.004**	**0.003**	**-0.012**	**-0.025**	**0.008**
	(0.011)	**(0.019)**	**(0.014)**	**(0.019)**	**(0.014)**
Girl bias (standardized beta)	-0.002	0.002	-0.007	-0.015	0.005
Observations	52,737	19,795	32,942	19,815	32,922
R-squared	0.329	0.357	0.260	0.340	0.269
Panel B: Black, only firstborn boys					
Girl bias	**0.014**	**0.020**	**-0.079**	**0.114**	**-0.027**
	(0.049)	**(0.058)**	**(0.118)**	**(0.076)**	**(0.075)**
Girl bias (standardized beta)	0.008	0.010	-0.046	0.059	-0.015
Observations	5,002	3,995	1,007	2,506	2,496
R-squared	0.478	0.488	0.710	0.572	0.545

Notes. This table reports OLS estimates, with robust standard errors clustered at the student and school level. 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 none of the children has unknown father. From these families we keep students enrolled in grades 6th to 10th for whom we have a mathematics test score. In this table we look only at firstborn male students, excluding only children. Sample statistics for this sample are reported in Appendix Table A2, Panel B. Columns (1) to (5) are defined in the same way as in Table 2. "Girl bias" is a dummy equal to 1 if the last born child in the family is a girl, and all the older children are boys, equal to 0 otherwise. The dependent variable and the controls are identical to the ones described in Table 2. All columns include year FE, grade FE, school FE. ***, **, and * indicate significance respectively at the 1%, 5%, and 10% levels.

Online Appendix for

THE FAMILY ORIGIN OF THE MATH GENDER GAP IS A WHITE AFFLUENT PHENOMENON

By GAIA DOSSI, DAVID FIGLIO, PAOLA GIULIANO, PAOLA SAPIENZA

DECEMBER 2020

Table A1 – Sample Statistics, Florida Department of Education

	All			White			Black		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.	N
Math score	0.161	0.929	703,654	0.340	0.863	489,903	-0.406	0.923	154,253
Female	0.504	0.500	703,654	0.497	0.500	489,903	0.524	0.499	154,253
White	0.696	0.460	703,654	1.000	0.000	489,903	0.000	0.000	154,253
Black	0.219	0.414	703,654	0.000	0.000	489,903	1.000	0.000	154,253
Hispanic	0.043	0.203	703,654	0.000	0.000	489,903	0.000	0.000	154,253
Asian	0.002	0.048	703,654	0.000	0.000	489,903	0.000	0.000	154,253
Other race	0.038	0.192	703,654	0.000	0.000	489,903	0.000	0.000	154,253
Median income zipcode of birth*100,000 (USD)	0.458	0.134	703,654	0.484	0.132	489,903	0.375	0.105	154,253
Free Lunch	0.429	0.495	703,654	0.308	0.462	489,903	0.782	0.413	154,253
Special Education	0.107	0.309	703,654	0.100	0.300	489,903	0.130	0.337	154,253
Age (in months)	153.119	14.616	703,654	152.928	14.624	489,903	154.176	14.691	154,253
Mother married at birth	0.596	0.491	703,654	0.738	0.440	489,903	0.185	0.388	154,253
Mother age at birth of 1st child	24.891	6.007	703,654	26.190	5.845	489,903	21.329	5.065	154,253
Mother high school dropout	0.190	0.393	703,654	0.139	0.346	489,903	0.337	0.473	154,253
Mother graduated high school	0.356	0.479	703,654	0.346	0.476	489,903	0.387	0.487	154,253
Mother attended some college	0.250	0.433	703,654	0.266	0.442	489,903	0.195	0.396	154,253
Mother graduated from college	0.204	0.403	703,654	0.249	0.432	489,903	0.081	0.273	154,253

Notes. The table reports descriptive statistics for the Florida sample used in Table 1 of the main text. The unit of observation is a student-year. The sample includes all students born in Florida between 1994 and 2002, attending a Florida public school grade 6 to 10 in years 2002 to 2011, and for whom we have a score in mathematics. Columns (1) to (3) report mean, st. deviation and sample size for the full sample of students. Columns (4) to (6), and Columns (7) to (9) report the same statistics respectively for the subsample of White students, and for the one of Black students. "Math score" measures students' Florida Comprehensive Assessment Test math score in a given grade (standardized with mean 0 and standard deviation 1 over the population for a given grade and year). The race dummies are taken from the FLODE school records. "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. "Special Education" is a dummy equal to 1 if the student is enrolled in the special education program in the given academic year. "Age in months" is the student's age at the beginning of the academic year. "Mother married at birth" is a dummy variable equal to 1 if the mother was married when the child was born. "Mother age at birth of 1st child" is the age (in years) of the mother when she gave birth to her first child. "Mother graduated high school", "Mother attended some college", "Mother graduated from college" are dummy variables with excluded category "Mother is a high school dropout".

Table A2- Sample Statistics, Florida Department of Education

	White			Black		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Sample: Firstborn females	Mean	St. Dev.	N	Mean	St. Dev.	N
Math score	0.482	0.799	50,402	-0.154	0.836	5,455
Boy bias	0.503	0.500	50,402	0.471	0.499	5,455
Median income zipcode of birth*100,000 (USD)	0.499	0.137	50,402	0.396	0.110	5,455
Free Lunch	0.211	0.408	50,402	0.612	0.487	5,455
Special Education	0.058	0.234	50,402	0.071	0.258	5,455
Age (in months)	157.718	16.122	50,402	158.867	16.260	5,455
Mother married at birth	0.875	0.330	50,402	0.480	0.500	5,455
Mother age at birth of 1st child	27.702	5.219	50,402	24.453	5.311	5,455
Mother graduated high school	0.301	0.459	50,402	0.371	0.483	5,455
Mother attended some college	0.283	0.451	50,402	0.285	0.451	5,455
Mother graduated from college	0.339	0.473	50,402	0.202	0.401	5,455
Panel B. Sample: Firstborn males						
Math score	0.527	0.865	52,737	-0.268	0.945	5,002
Girl bias	0.476	0.499	52,737	0.500	0.500	5,002
Median income zipcode of birth*100,000 (USD)	0.502	0.139	52,737	0.391	0.110	5,002
Free Lunch	0.209	0.407	52,737	0.615	0.487	5,002
Special Education	0.119	0.323	52,737	0.150	0.357	5,002
Age (in months)	158.176	16.033	52,737	159.978	16.273	5,002
Mother married at birth	0.880	0.325	52,737	0.478	0.500	5,002
Mother age at birth of 1st child	27.857	5.207	52,737	24.345	5.428	5,002
Mother graduated high school	0.302	0.459	52,737	0.358	0.479	5,002
Mother attended some college	0.273	0.446	52,737	0.298	0.457	5,002
Mother graduated from college	0.351	0.477	52,737	0.201	0.401	5,002

Notes. The table reports descriptive statistics for the Florida sample used in Table 2 and Table 3. The unit of observation is a student-year. The observations shown in this table are all subsamples of the sample shown in Appendix Table 1. In Panel A, Columns (1) to (3) we report mean, st. deviation and number of observations for the subsample of White firstborn females. This corresponds to the sample used in Table 2, Panel A, Column (1). In Panel A, Columns (4) to (6) we report mean, st. deviation and number of observations for the subsample of Black firstborn females. This corresponds to the sample used in Table 2, Panel B, Column (1). In Panel B, Columns (1) to (3) we report mean, st. deviation and number of observations for the subsample of White firstborn males. This corresponds to the sample used in Table 3, Panel A, Column (1). In Panel B, Columns (4) to (6) we report mean, st. deviation and number of observations for the subsample of Black firstborn males. This corresponds to the sample used in Table 3, Panel B, Column (1). "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. "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. The rest of the variables is defined as in Appendix Table 1. In our definition of firstborns we always exclude only children.

Description of the Variables

Name of the variable	*Description*	*Source (and when possible and useful name of the raw variable)*
Math score	Development scale score in the Mathematics section of the FCAT. 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 in the sample of girl and boys of families for which we observe completed fertility, 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	A dummy equal to 1 if the last born in the family is a girl, and all the older children are boys, 0 otherwise.	Source: birth certificate, FLDOE Created using raw variables: GENDER_CD
Female	A dummy equal to 1 if the student is a girl, 0 otherwise.	Source: FLDOE Created using raw variables: GENDER_CD
Race dummies (White, Black, Hispanic, Asian, Other race)	A set of dummies equal to 1 if the student is of that ethnicity, 0 otherwise.	Source: FLDOE Created using raw variables: RACIAL_ETHNIC_CD
Median income in zip code of birth, (100,000 of $)	The zipcode at time of birth (provided by the birth certificates) is matched with median zipcode 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 0 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
Special Education	A dummy variable equal to 1 if the variable if the student is enrolled in the special education program, 0 otherwise. Gifted students are classified as 0.	Source: FLDOE Created using raw variables: PRIMARY_EXCPT_IND
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
Mother age at first birth	Age of the mother when the mother's first child was born. The variable was calculated using mother's year and month of birth, and child's year and month of birth.	Source: FLDOE, birth certificate STUDENT_BIRTH_MONTH, STUDENT_BIRTH_YEAR

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, 0 otherwise.	Source: FLDOE Created using raw variables: LUNCH_STATUS