The Gender Math Gap: Is It Growing?

Kitae Sohn[*]

The Department of Economics, University at Albany, SUNY,

1400 Washington Ave. Albany, NY, 12222, USA

Abstract

The gender math gap in the USA has been declining and the gap, if any, is believed to appear in adolescence. Recently it is reported not only that the gap is observed in childhood but also that the gap is growing, as the grade advances. Using the Early Childhood Longitudinal Study, I estimate the counterfactual distribution of girls' math scores and measure the gap for each quantile. I find that the gap is not growing but, as girls age, more girls experience a similar size of the gap starting from girls at the top followed by girls in the lower parts of the distribution of math skills. The results suggest that more targeted remediation would be more efficient to close the gender math gap.

JEL classification: J24; J31; I21

Keywords: Gender math gap; Counterfactual distribution; Early Childhood Longitudinal Study

---

[*] Tel.: 1 518 489 0357
E-mail address: ks126361@albany.edu

1. Introduction

Recently Fryer and Levitt (2006) incidentally find that the math gap in the USA not only exists in the early elementary school years but also grows. Although this is not the first finding of the math gap in childhood (e.g. Anastasi, 1958; Maccoby, 1966; Geary, 1994), their finding is surprising in that the gap is reported to be closing (Campbell et al., 1999). Moreover psychologists generally agree that, if the gap ever exists, it appears in adolescence at the earliest (Kimball, 1989; Hyde, Fennema, and Lamon, 1990; Hedges and Nowell, 1995; Leahey and Guo, 2001).

If their finding is true, it has important implications for the gender wage gap, because the importance of the math gap is recognized to explain the wage gap (Daymont and Andrisani 1984; Paglin and Rufolo, 1990; Brown and Corcoran, 1997).[1] Also, since much evidence substantiates “learning begets learning” and little malleability of cognitive skills even in adolescence (see Heckman and Masterov, 2007 and Heckman and his colleagues’ works in the references), the early appearance of the math gap pose a serious challenge to the policy-makers in education.

In this paper, I attempt to answer a simple question: whether the gap is actually growing or the gap is similar for each girl, but more girls experience the gap. For example, consider the cases illustrated in Table 1. In Case A, both girls fall behind in each period by the same size compared to boys with the corresponding ability. The average gap increases between the two periods. Case B also describes the widening gap of the same amount as in Panel A on average, but the gap exists initially only for the

---

[1] Angle and Wissman (1981) and Filer (1983) find little relationship between math skills and earnings. Daymont and Adrisani (1984), however, contend that the former misinterpret the results and the sample of the latter is not representative.

high-ability girl and it does not grow in the next period for her. The average gap increases only because the low ability girl, too, experiences the gap of the same size in the second period.

Depending on the situation, the policy implication is different. If the policy goal is closing the gap, more resources need to be allocated evenly each period in Case A. On the other hand, in Case B, it is more efficient to target the resources for the high ability girl in Period 1 and for the low ability girl in Period 2. With the same amount of resources, if the resource allocation suitable for Case A is adopted for Case B, the goal would likely to fail. Even if the extra resources help the low ability girl outperform her male counterpart, the issue of equity still remains.

By estimating the counterfactual distribution of girls' math scores, I find that the evolution of the math gap is more similar to Case B. Girls in the top distribution initially falls behind boys, followed by girls in the lower parts of the distribution. Moreover, girls at the top recover the initial loss, so overall the math gap stabilizes at 0.2 SD in the entire distribution. The results suggest the reconsideration of the current policy in education, which is in much favor of students at the bottom.

The remainder of the paper is structured as follows: Section 2 describes the data, Section 3 discusses the empirical approach, Section 4 reports the results, and Section 5 concludes.

2. Data

The ECLS starts following the nationally representative kindergarten cohort of over 20,000 children in 1998. The data on the full sample was obtained in the fall and spring of kindergarten, spring of first grade, spring of third grade, and spring of fifth grade. At each survey, children took a battery of standardized tests including the math test. Since there is no natural unit of math skills, I standard normalized item response theory (IRT) math scores.

The data contain a wide range of information collected from parents, teachers, and school administrators. Fryer and Levitt (2004) succeed in eliminating black-white test score gap in kindergarten with a small number of covariates and demonstrate that adding more variables improves little the goodness of fit. Later Fryer and Levitt (2006) conduct the same analysis with the longer sample period and, again, find that adding more variables in addition to their parsimonious variables affect little their results. The little improvement results from the fact that the additional variables are highly correlated with the parsimonious variables, their coefficients are not statistically significant, or if significant, the size is modest. So, in the following analyses, I use their parsimonious variables only and make the specifications more efficient.

I drop observations without math scores or race. I combine race specified Hispanic and non-specified Hispanic. Also the following races are reclassified into “other race”: native Hawaiian, other Pacific Islander, American Indian or Alaska native, and non-Hispanic multi-race. The age at the kindergarten entry (in months) is used as reported. The birth weight (in ounces) and the WIC participation are updated if not reported in the previous interviews. The nonresident biological mother’s age at first birth is used, if the current mother's age at first birth is not available.[2] I treat all observations marked with "refused," "don't know," "not ascertained," and "not applicable" (unless updated) as missing observations. Finally I apply panel weights (C1_6FP0) for all the following analyses.[3]

Table 2 presents the descriptive statistics of math scores and independent variables by gender and the differences. As Fryer and Levitt (2006) point out, girls continue to lag behind boys in math. There is no detectable gap in kindergarten, but a small gap arises in first grade. In third grade, the gap abruptly increases, followed by another small increase in fifth grade. By fifth grade, the gap reaches 0.2 SD.

Another (un)interesting finding is that boys and girls show observably little difference besides the scores. Boys and girls are equally born to households regardless of the socioeconomic status. They have an equivalent number of books in every grade except third grade and both enter kindergarten at the same age. Boys are born slightly heavier, but the difference is small. And the same pattern emerges for the mothers' age at first birth as well as the WIC participation rate. The statistics seem to be reasonable for the USA. Unless sex selective abortion is widely practiced, the sex of children should be randomly distributed irrespective of the socioeconomic status and mothers' characteristics.[4] If boys and girls are equally valuable, they should access to the same learning environment, although the content of the environment may be different. One can

---

[2] Note that I do not recode the mother's age at first birth into under 20, 20s, and 30 or above as Fryer and Levitt (2004, 2006) do, because the cut-points seem to be arbitrary. Using their way, however, does not affect the results.

[3] It is meaningless to enter the dummy indicating missing observations for each variable as Fryer and Levitt (2004, 2006) do, because zero panel weights are assigned to the missing observations, except a few observations.

[4] There is evidence that girls (boys) are more likely born to mothers in bad (good) condition (Trivers and Willard, 1973; Almond and Edlund, 2007). But the size is not substantive for this study. For example, compared to a mother with some college, a mother without a high school degree has a lower chance of having a boy only by 0.6 percent.

anticipate from the statistics that the math gap results from the effects of the characteristics not from the characteristics per se. In fact, this is the case.

3. Methods

The methodological intuition is a “what if” question: what if girls receive a treatment to be boys, if so how much the math gap would be. The first question can be addressed in the context of the average treatment effects on the treated (ATT) and the Oaxaca-Blinder decomposition help answer the second question (Oaxaca, 1973; Blinder, 1973). Melly (2005, 2006) illustrates that the decomposition method contains the concept of the ATT in itself and that the decomposition can be extended to estimate counterfactual distributions using the quantile regression model.

The decomposition was initially developed to decompose the gender wage gap into explained and unexplained parts by observable characteristics. But it has also been applied to assess the gap of test scores between countries (e.g. McEwan, and Marshall, 2004; Ammermueller, 2007) or schools (e.g. Krieg and Storer, 2006). To my best knowledge, the method has not been applied for the gender math gap covering the entire distribution of math skills.

Suppose that math scores, Y, are linearly estimated with a vector of the parsimonious variables, X, and a female dummy, F, for student i in grade g. Then the econometric specification would be

$$Y_{ig} = X_{ig}\beta_g + \gamma F_i + u_{ig} \tag{1}$$

where β and γ are the coefficients on the corresponding independent variables and u_{ig} is a random error. The decomposition starts with the two separate regressions by gender as follows.

$$Y_{ig}^{t} = X_{ig}^{t}\beta_{g}^{t} + u_{ig}^{t}, t = m, f, \tag{2}$$

where m and f indicate boys and girls, respectively. From the regressions, the math gap at the mean is estimated as

$$\overline{Y^{f}} - \overline{Y^{m}} = \left[\overline{X^{f}}\,\hat{\beta}^{f} - \overline{X^{f}}\,\hat{\beta}^{m}\right] + \left[\overline{X^{f}}\,\hat{\beta}^{m} - \overline{X^{m}}\,\hat{\beta}^{m}\right],$$ [5]

where $\overline{Y^{t}} = n_t^{-1}\sum_{i:T_i=t} Y_i$, $\overline{X^{t}} = n_t^{-1}\sum_{i:T_i=t} X_i$ with n_t number of observations for t and gender T_i. The first and second brackets contain the effects of coefficients and the effects of characteristics, respectively. So the former measures the gap between the actual female test score and their counterfactual test score, where the counterfactual represents what if girls have the effects of boys. On the other hand, the second bracket estimates the gap resulting from the different characteristics, when the male effects are considered as the norm.

As Melly (2006) correctly recognizes, the first bracket is conceptually same as the ATT. In the context of the ATT, the usual index number problem, that plagues the Oaxaca-Blinder decomposition, causes little concern. The problem occurs, because there is no a priori reason to choose β^{m} over β^{f} as the norm. But the question of this study is "what if girls are in boys' situation with their own characteristics?" Hence the question inevitably poses β^{m} as the norm. Moreover, as seen earlier and will be demonstrated

---

[5] Subscripts i and g are suppressed to simplify the notation.

later, $\overline{X^m} \approx \overline{X^f}$, so the second bracket essentially becomes zero. The gap solely comes from the different coefficients and the question is answered just as would be done by the ATT only.

The main methodology of this study is decomposing the math scores over the whole distribution not just at the mean as the conventional Oaxaca-Blinder decomposition does. Some attempts have been made for this purpose, for example, by Juhn, Murphy, and Pierce (1993), DiNardo, Fortin, and Lemieux (1996), Lemieux (2002), Machado and Mata (2005). However, I adopt Melly's (2006) methodology for this paper. His methodology accounts for heteroscedasticity (the problem with Juhn et al. (1993)), overcomes "the curse of dimension" (the problem with DiNardo et al. (1996) and Lemieux (2002)) and is more efficient (superior to Machado and Mata (2005)) (see Melly, 2005 for more explanation).

Following his notation, let $F_Y(q)$ and $f_Y(q)$ be the cumulative distribution function of random variable Y at q and its corresponding probability density function, respectively. Then, the inverse of $F_Y(\cdot)$, $F_Y^{-1}(\theta)$, is the quantile function evaluated at $0<\theta<1$. If $F_Y^{-1}(\theta)$ is linearly determined by X_i,[6] then

$$F_{Y(t)}^{-1}(\tau \mid X_i) = X_i \beta_t(\tau)$$

where $0<\tau<1$. According to Koenker and Bassett (1978),

$$\hat{\beta}_t(\tau) = \arg\min_{b \in \Re} n_t^{-1} \sum_{i:T_i=t} \rho_t(Y_i - X_i b),$$

where $\rho_\tau(z) = z(\tau - 1(z \le 0))$ with $1(\cdot)$ being the indicator function.

---

[6] g is suppressed.

The goal is to obtain the unconditional quantiles, which are derived from $F^{-1}_{Y(t)}(\cdot)$, to estimate the counterfactual scores of girls. The procedure would be straightforward, if the estimated $F^{-1}_{Y(t)}(\cdot \mid X_i)$ is monotonous:

Step 1. Estimate $F^{-1}_{Y(t)}(\cdot \mid X_i)$ by (2)

Step 2. $F^{-1}_{Y(t)}(\cdot \mid X_i)$ → Invert → $F_{Y(t)}(\cdot \mid X_i)$

Step 3. $F_{Y(t)}(\cdot \mid X_i)$ → Integrate over X_i → $F_{Y(t)}(\cdot)$

Step 4. $F_{Y(t)}(\cdot)$ → Invert → $F^{-1}_{Y(t)}(\cdot)$

However, the estimated $F^{-1}_{Y(t)}(\cdot \mid X_i)$ is not necessarily monotonous. This quantile crossing prevents Step 2. Since $F_{Y(t)}(\cdot \mid X_i)$ has the following property,

$$F_{Y(t)}(q \mid X_i) = \int_0^1 1\left(F^{-1}_{Y(t)}(\tau \mid X_i) \le q\right) d\tau = \int_0^1 1\left(X_i \beta_i(\tau) \le q\right) d\tau ,$$

one can estimate $F_{Y(t)}(\cdot \mid X_i)$ by

$$\hat{F}_{Y(t)}(q \mid X_i) = \int_0^1 1\left(X_i \hat{\beta}_i(\tau) \le q \right) d\tau = \sum_{j=1}^{J} \left(\tau_j - \tau_{j-1}\right) 1\left(X_i \hat{\beta}_i(\tau_j) \le q \right),$$

where $\hat{\beta}_i(\tau_j)$ prevails in the interval between τ_{j-1} and τ_j with τ_j being the point where the solution changes (Portnoy, 1991). This step corresponds to Step 2. From this, the estimation of $F_{Y(t)}(\cdot)$ can be obtained by

$$\hat{F}_{Y(t)}(q \mid T = t) = \int \hat{F}_{Y(t)}(q \mid x) dF_X(x \mid T = t) = n_t^{-1} \sum_{i:T_i = t} \hat{F}_{Y(t)}(q \mid X_i) .$$

This is equivalent to Step 3. Finally, each estimated quantile q at $\tau = \theta$ is given by

$$\hat{q}_t(\theta) = \inf\left\{ q : n_t^{-1} \sum_{i:T_i=t} \hat{F}_{Y(t)}(q \mid X_i) \geq \theta \right\},$$

and the counterfactual quantile becomes

$$\hat{q}_c(\theta) = \inf\left\{ q : n_t^{-1} \sum_{i:T_i=f} \hat{F}_{Y(f)}(q \mid X_i) \geq \theta \right\}.$$

Hence, the quantile version of the Oaxaca-Blinder decomposition can be expressed as

$$\hat{q}_f(\theta) - \hat{q}_m(\theta) = \left[\hat{q}_f(\theta) - \hat{q}_c(\theta) \right] + \left[\hat{q}_c(\theta) - \hat{q}_m(\theta) \right]. \tag{3}$$

The interpretation is same as (2): the first and second brackets represent the gaps due to the difference in the coefficients and characteristics, respectively, at $\tau = \theta$.

Melly (2006) proves that $\hat{q}_t$ and $\hat{q}_c$ are consistent and asymptotically normally distributed with the traditional assumptions for the quantile regression models and some additional (and trivial for this study) assumptions.

4. Results

4.1. The math gap at the mean

I start with the presentation that the math gap grows at the mean, as reported by Fryer and Levitt (2006), and the phenomenon is robust to different specifications. Panel A in Table 3 restates the widening math gap. The raw difference is same as the math difference in the descriptive statistics. The between school difference[7] is estimated with (1), whereas school dummies are added to (1) to estimate the within school difference.[8] Note that controlling covariates does not change the level of the math gap in each grade and the increasing rate of the gap across the grades. This implies that boys and girls are orthogonal to the covariates, i.e. the two sexes are randomly distributed. The within school difference strengthens the interpretation. The math gap in the same schools does not deviate much from the raw difference or the between school difference. Although the math gap already exists in kindergarten, the gap is modest. Inferring that girls attend better kindergarten schools seems to stretch the outcome too far.

From the results, it is not hard to conjecture the outcomes of the Oaxaca-Blinder decomposition: the characteristics part should be zero. Panel B confirms the conjecture.[9] The columns with "male" present the results of (2), when the coefficients of boys are taken as the norm. This is the counterfactual argument that this paper is seeking.[10] Although not a critical point for this study, since the index number problem is such an important issue in the decomposition, I check the problem with the coefficients estimated with the pooled sample suggested by Neumark (1988). As shown in the columns of "pooled," the math gap is, again, entirely explained by the effects of the coefficients. The results do not change with other schemes, e.g. Reimers' (1983) method or taking the female coefficients as the norm (available upon request).

4.2. The math gap in the distribution

---

[7] The regression results are reported in Table A-1 in the appendix.
[8] The regression results are reported in Table A-2 in the appendix.
[9] Standard errors are calculated as Jann (2005) proposes.
[10] The regression results are reported in Table A-3 in the appendix.

The results so far give the impression that the average math gap increases and the growing gap is robust to different specifications. And yet, the average conceals more than reveals. When the decomposition is extended to the quantile model, it becomes clear whether the situation is similar to Case A or Case B in Table 1. Using (3), I depict the evolution of the math gap. First, it is obvious that the characteristics have no role in the entire distribution as well as at the mean. Hence the random birth and equivalent treatment applies to every level of math skills as well as at the mean.

More important finding is that girls at the top start falling behind already in the spring of kindergarten. By contrary, girls at the bottom show a hint of outperforming boys in first grade. As girls at the top start recovering, however, girls in other parts of the distribution lose ground compared to boys in each grade. By fifth grade, the whole distribution of girls is below that of boys by about 0.2 SD. And yet, the gap in each quantile is stable around 0.2 SD. The increasing math gap is, in fact, not due to the shift of the entire distribution of girls’ math scores in each grade. The reason for the seemingly growing math gap is the increasing numbers of girls who have lower scores than boys with the gap of a similar size for each quantile.

At this point, it is not certain how the gap would evolve beyond the childhood of the sample. Considering the evolution of the math gap across the grades, it appears that girls at the top would recover or even become at the par with boys, whereas girls at the bottom lose further ground. If this would be the case, the average gap would not increase. But, under the still surface of the unchanging gap, thorny issues, e.g. equity in education or earnings, would be growing. Although this is a speculation, it is worth considering for the gender math gap in the short term and in earnings in the long term.

5. Conclusion

The gender math gap is believed to be declining or absent for the last decades. And yet, the gap reappears unexpectedly in childhood. By using the ECLS, I estimate the counterfactual distribution of girls' math scores and demonstrate that the gap is not growing but more girls lag behind in each grade with a similar size of the gap.

By denying the growing gap, I do not discount the importance that the size of 0.2 SD implies. Actually, it has a dramatic effect at the upper tail of the distribution. For example, if the test scores are equally distributed as standard normal for both girls and boys and the variances are same,[11] boys would be present more relative to girls by 53.8 percent and 73.6 percent at the top 5 and 1 percent of the distribution, respectively.

A more acute point of the results is, however, that the current policy in education seems to pay disproportionate attention to students at the bottom, as demonstrated by the Head Start, Title 1, or the No Child Left Behind. As demonstrated above, it is girls at the top who first lose ground compared to boys in math. If the dynamic formation of cognitive skills is true ("learning begets learning"), it is possible that they will not catch up to boys in adolescence and adulthood. Also, considering the link of the math skills to wages, more thought should be given to closing their gap as early as possible.

It is out of question that girls at the bottom as well as in the middle of the distribution also should get help to close the math gap. The subtle issue is when they

---

[11] The latter assumption yield a conservative ratio of boys to girls at the top, because it is widely reported that the variance of female math scores is smaller the variance of male math scores. So even if both means are same, there are more boys than girls at the upper tail (Hedges and Nowell, 1995)

receive the help. More resources should not harm. But just providing more resources to every one is not an efficient way to close the gap. The resources should be targeted to girls who need most. The results suggest that girls at the top receive the resources first and the resources be shifted gradually to girls in the lower parts of the distribution.

This paper leaves many questions unanswered. Most of all, it is not clear why girls at the top fall behind boys first and what helps them recover. Also it is hard to know why girls at the bottom, who, even lead boys in first grade, cannot keep the lead and eventually lose. Since the literature on the gender math gap exclusively focuses on the average gap, it does not help much to understand the conflicting and heterogeneous pattern across the distribution. This paper does not provide the answers to the important questions, but it opens a new path to searching for the answers.

Appendix

Table A-1 The estimated female-male math score gap

	Spring Kindergarten	Fall Kindergarten	Spring 1st Grade	Spring 3rd Grade	Spring 5th Grade
FEMALE	-0.021 (0.030)	-0.053 (0.033)	-0.067 (0.034)†	-0.168 (0.035)‡	-0.185 (0.034)‡
BLACK	-0.064 (0.047)	-0.176 (0.054)‡	-0.327 (0.054)‡	-0.366 (0.062)‡	-0.485 (0.061)‡
HISPANIC	-0.057 (0.043)	-0.021 (0.048)	-0.073 (0.047)	-0.020 (0.055)	0.031 (0.050)
ASIAN	0.241 (0.071)‡	0.194 (0.089)†	-0.003 (0.088)	0.042 (0.092)	0.186 (0.080)‡
OTHER RACE	-0.092 (0.070)	-0.075 (0.078)	-0.143 (0.070)‡	-0.154 (0.072)†	-0.105 (0.084)
SES	0.246 (0.032)‡	0.236 (0.030)‡	0.278 (0.029)‡	0.311 (0.028)‡	0.329 (0.029)‡
BOOK	0.006 (0.001)‡	0.006 (0.001)‡	0.001 (<0.000)‡	<0.000 (<0.000)‡	<0.000 (<0.000)‡
BOOK^2 (X 1000)	-0.019 (0.005)‡	-0.021 (0.005)‡	<-0.000 (<0.000)‡	<-0.000 (<0.000)‡	<-0.000 (<0.000)‡
ENTRY AGE	0.054 (0.004)‡	0.054 (0.005)‡	0.042 (0.004)‡	0.029 (0.004)‡	0.014 (0.004)‡
BIRTH WEIGHT	0.022 (0.007)‡	0.027 (0.008)‡	0.029 (0.008)‡	0.027 (0.008)‡	0.028 (0.008)‡
AGE AT 1st BIRTH (X 10)	0.025 (0.003)‡	0.020 (0.004)‡	0.017 (0.003)‡	0.024 (0.004)‡	0.024 (0.004)‡
WIC	-0.184 (0.039)‡	-0.157 (0.045)‡	-0.141 (0.043)‡	-0.118 (0.047)†	-0.079 (0.044)*
CON	-4.454 (0.298)‡	-4.404 (0.345)‡	-3.300 (0.331)‡	-2.530 (0.316)‡	-1.619 (0.320)‡
R2	0.277	0.242	0.220	0.253	0.268
N	6889	6889	6942	6905	6945

Notes: The dependent variable is standard normalized IRT math scores. The independent variables are explained in the note of Table 2. Panel weights are applied. ‡: p-value <0.01; †: p-value <0.05; *: p-value <0.10.

Table A-2 The estimated female-male math score gap with school fixed effects

	Spring Kindergarten	Fall Kindergarten	Spring 1st Grade	Spring 3rd Grade	Spring 5th Grade
FEMALE	-0.061 (0.028)†	-0.069 (0.028)†	-0.090 (0.026)‡	-0.171 (0.028)‡	-0.166 (0.029)‡
BLACK	-0.191 (0.060)‡	-0.241 (0.066)‡	-0.346 (0.066)‡	-0.421 (0.071)‡	-0.501 (0.090)‡
HISPANIC	-0.055 (0.049)	-0.034 (0.055)	-0.110 (0.050)†	-0.092 (0.053)*	-0.072 (0.063)
ASIAN	0.275 (0.082)‡	0.270 (0.093)‡	0.123 (0.088)	0.112 (0.095)	0.149 (0.077)*
OTHER RACE	-0.016 (0.071)	0.002 (0.091)	-0.009 (0.084)	0.004 (0.096)	-0.052 (0.135)
SES	0.218 (0.025)‡	0.225 (0.027)‡	0.252 (0.025)‡	0.253 (0.025)‡	0.260 (0.026)‡
BOOK	0.006 (0.001)‡	0.006 (0.001)‡	0.001 (<0.000)‡	0.001 (<0.000)‡	0.000 (<0.000)†
BOOK^2 (X 1000)	-0.020 (0.004)‡	0.021 (0.004)‡	<-0.000 (<0.000)†	<-0.000 (<0.000)‡	<0.000 (<0.000)
ENTRY AGE	0.056 (0.004)‡	0.052 (0.004)‡	0.037 (0.004)‡	0.025 (0.004)‡	0.016 (0.004)‡
BIRTH WEIGHT	0.031 (0.006)‡	0.032 (0.007)‡	0.039 (0.007)‡	0.036 (0.006)‡	0.023 (0.006)‡
AGE AT 1st BIRTH (X 10)	0.017 (0.003)‡	0.014 (0.003)‡	0.011 (0.003)‡	0.016 (0.003)‡	0.015 (0.003)‡
WIC	-0.109 (0.036)‡	-0.122 (0.040)‡	-0.103 (0.037)‡	-0.102 (0.039)‡	-0.131 (0.043)‡
CON	-4.493 (0.261)‡	-4.175 (0.282)‡	-2.963 (0.278)‡	-2.180 (0.273)‡	-1.401 (0.294)‡
adj. R2	0.401	0.392	0.433	0.511	0.550
N	6889	6888	6942	6905	6945

Notes: The dependent variable is standard normalized IRT math scores. The independent variables are explained in the note of Table 2. The coefficients on school dummies are omitted. Panel weights are applied. ‡: p-value <0.01; †: p-value <0.05; *: p-value <0.10.

Table A-3 The OLS estimation of math score by gender

	Fall Kindergarten		Spring Kindergarten		Spring 1st Grade		Spring 3rd Grade		Spring 5th Grade	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
BLACK	-0.044	-0.071	-0.221	-0.121	-0.390	-0.237	-0.445	-0.274	-0.510	-0.451
	(0.073)	(0.058)	(0.079)‡	(0.070)*	(0.083)‡	(0.067)‡	(0.087)‡	(0.083)‡	(0.087)‡	(0.084)‡
HISPANIC	-0.085	-0.028	-0.073	0.028	-0.126	-0.011	-0.101	0.068	-0.005	0.063
	(0.065)	(0.056)	(0.071)	(0.063)	(0.072)*	(0.057)	(0.085)	(0.067)	(0.076)	(0.064)
ASIAN	0.408	0.138	0.367	0.090	0.062	-0.030	0.134	0.002	0.304	0.119
	(0.110)‡	(0.086)	(0.125)‡	(0.112)	(0.141)	(0.110)	(0.110)	(0.127)	(0.087)‡	(0.116)
OTHER RACE	-0.031	-0.144	0.018	-0.162	-0.034	-0.241	-0.110	-0.186	-0.093	-0.110
	(0.097)	(0.095)	(0.119)	(0.092)*	(0.107)	(0.079)‡	(0.122)	(0.078)†	(0.145)	(0.089)
SES	0.244	0.245	0.236	0.229	0.309	0.229	0.317	0.298	0.331	0.320
	(0.047)‡	(0.041)‡	(0.043)‡	(0.041)‡	(0.043)‡	(0.038)‡	(0.038)‡	(0.039)‡	(0.039)‡	(0.041)‡
BOOK	0.007	0.005	0.005	0.007	0.001	0.001	0.001	<0.000	<0.000	<0.000
	(0.002)‡	(0.001)‡	(0.002)‡	(0.002)‡	(<0.000)‡	(<0.000)‡	(<0.000)‡	(<0.000)†	(<0.000)†	(<0.000)†
BOOK^2 (X 1000)	-0.022	-0.017	-0.015	-0.027	<-0.000	<-0.000	<-0.000	<-0.000	<-0.000	<-0.000
	(0.007)‡	(0.006)†	(0.008)*	(0.007)‡	(<0.000)‡	(<0.000)‡	(<0.000)‡	(<0.000)‡	(<0.000)	(<0.000)‡
ENTRY AGE	0.060	0.049	0.059	0.049	0.046	0.036	0.031	0.026	0.013	0.015
	(0.006)‡	(0.006)‡	(0.007)‡	(0.006)‡	(0.007)‡	(0.006)‡	(0.006)‡	(0.005)‡	(0.006)†	(0.006)†
BIRTH WEIGHT	0.013	0.031	0.010	0.048	0.015	0.047	0.012	0.045	0.015	0.046
	(0.010)	(0.009)‡	(0.012)	(0.010)‡	(0.012)	(0.001)‡	(0.011)	(0.012)‡	(0.011)	(0.013)‡
AGE AT 1st BIRTH	0.020	0.029	0.020	0.019	0.018	0.016	0.019	0.028	0.027	0.022
	0(0.005)‡	(0.004)‡	(0.005)‡	(0.005)‡	(0.005)‡	(0.004)‡	(0.005)‡	(0.005)‡	(0.005)‡	(0.005)‡
WIC	-0.189	-0.182	-0.148	-0.171	-0.107	-0.182	-0.138	-0.107	-0.032	-0.136
	(0.061)‡	(0.047)‡	(0.071)†	(0.052)‡	(0.067)	(0.051)‡	(0.071)*	(0.059)*	(0.066)	(0.057)‡
CON	-4.635	-4.312	-4.490	-4.373	-3.438	-3.171	-2.373	-2.815	-1.498	-1.975
	(0.435)‡	(0.404)‡	(0.548)‡	(0.411)‡	(0.497)‡	(0.429)‡	(0.464)‡	(0.419)‡	(0.438)‡	(0.459)‡
R2	0.261	0.303	0.232	0.264	0.226	0.224	3442	3463	3464	3481
N	3427	3462	3427	3462	3463	3479	0.249	0.252	0.252	0.272

Notes: The dependent variable is standard normalized IRT math scores. The independent variables are explained in the note of Table 2. Panel weights are applied. ‡: p-value <0.01; †: p-value <0.05; *: p-value <0.10.

References

Almond, D., Edlund, L., 2007. Trivers-Willard at birth and one year: Evidence from US natality data 1983-2001. Proceedings of the Royal Society of London B 274 (1624), 2491-2496.

Ammermueller, A., 2007. PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany. Empirical Economics 33 (2), 263-287.

Anastasi, A., 1958. Differential Psychology, 3rd ed. Macmillan, New York.

Angle, J., Wissman, D.A., 1981. Gender, college major, and earnings. Sociology of Education 54 (1), 25-33.

Blinder, A., 1973. Wage discrimination: reduced form and structural estimates. Journal of Human Resources 8 (4), 436-455.

Brown, C., Corcoran, M. 1997. Sex-based differences in school content and the male-female wage gap. Journal of Labor Economics 15 (3), 431-465.

Campbell, J.R., Hombo, C.M., Mazzeo, J., 1999. Trends in academic progress: Three decades of student performance. National Center for Education Statistics 2000-469.

Daymont, T.N., Andrisani, P.J., 1984, Job preferences, college major, and the gender gap in earnings. Journal of Human Resources 19 (3), 408-428.

DiNardo, J.E., Fortin, N.M., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. Econometrica 64 (5), 1001-1046.

Filer, R.K., 1983. Sexual differences in earnings: The role of individual personalities and tastes. Journal of Human Resources 18 (1), 82-99.

Fryer, R., Levitt, S., 2004. Understanding the black-white test score gap in the first two years of school. Review of Economics and Statistics 86 (2), 447-464.

Fryer, R.G., Levitt, S.D., 2006. The black-white test score gap through third grade. American Law and Economics Review 8 (2), 249-281.

Geary, D., 1994. Children’s Mathematical Development: Research and Practical Applications. American Psychological Association, Washington, D.C.

Heckman, J.J., Masterov, D.V., 2007. The productivity argument for investing in young children. NBER Working Paper 13016.

Hedge, L.V., Nowell, A., Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science 269 (5220), 41-45.

Hyde, J.S., Fennema, E., Lamon, S., 1990. Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin 107 (2), 139-155.

Jann, B., 2005. Standard errors for the Blinder-Oaxaca decomposition: http://repec.org/dsug2005/oaxaca-se-handout.pdf.

Juhn, C., Murphy, K.M., Pierce, B., 1993. Wage inequality and the rise in returns to skill. Journal of Political Economy 101(3), 410-442.

Kimball, M.M., 1989. A new perspective on women's math achievement. Psychological Bulletin. 105 (2), 198-214.

Krieg, J.M., Storer, P., 2006. How much do students matter? Applying the Oaxaca decomposition to explain determinants of adequate yearly progress. Contemporary Economic Policy 24 (4), 563-581.

Leahey, E., Guo, G., 2001. Gender differences in mathematical trajectories. Social Forces 80 (2), 713-732.

Lemieux, T., 2002. Decomposing changes in wage distributions: a unified approach. Canadian Journal of Economics 35 (4), 646-688.

Maccoby, E.E., 1966. The Differences in Intellectual Functioning. Stanford University Press, Stanford, CA.

Machado, J., Mata, J., 2005. Counterfactual decompositions of changes in wage distributions using quantile regression. Journal of Applied Econometrics 20 (4), 445-465.

McEwan, P.J., Marshall, J.H., 2004. Why does academic achievement vary across countries? Evidence from Cuba and Mexico. Education Economics 12 (3), 205-217.

Melly, B. 2006. Estimation of counterfactual distribution using quantile regression. Working Paper, University of St. Gallen.

Melly, B., 2005. Decomposition of differences in distribution using quantile regression. Labour Economics 12 (4), 577-590.

Neumark, D., 1988. Employers' discriminatory behavior and the estimation of wage discrimination. Journal of Human Resources 23 (3), 279-295.

Oaxaca, R., 1973. Male-female wage differentials in urban labor markets. International Economic Review 14 (3), 693-709.

Paglin, M., Rufolo, A.M., 1990. Heterogeneous human capital, occupational choice, and the male-female earnings. Journal of Labor Economics 8 (1), part 1, 123-144.

Portnoy, S., 1991. Asymptotic behavior of the number of regression quantile breakpoints. SIAM Journal of Scientific and Statistical Computing 12 (4), 867-883.

Reimers, C.W., 1983. Labor market discrimination against Hispanic and Black men. Review of Economics and Statistics 65 (4), 570-579

Trivers, R.L., Willard, D.E., 1973. Natural selection of parental ability to vary the sex ratio of offspring. Science 179 (4068), 90-92.

Table 1 Two possible cases of the evolution of the math gap

	Low Ability Girl	High Ability Girl	Average Gap
	Case A		
Period 1	0.1 SD	0.1 SD	0.1 SD
Period 2	0.2 SD	0.2 SD	0.2 SD
	Case B		
Period 1	0 SD	0.2 SD	0.1 SD
Period 2	0.2 SD	0.2 SD	0.2 SD

Table 2 Descriptive statistics

	Male	Female	Difference	N
MATH				
Fall Kindergarten	0.131	0.123	0.08	7147
Spring Kindergarten	0.135	0.089	0.046	
Spring 1st Grade	0.136	0.057	0.079†	
Spring 3rd Grade	0.181	-0.010	0.191‡	
Spring 5th Grade	0.145	-0.050	0.195‡	
SES				
Kindergarten	0.029	0.076	-0.047	7147
Spring 1st Grade	0.007	0.003	0.004	
Spring 3rd Grade	-0.033	-0.042	0.009	
Spring 5th Grade	-0.022	-0.010	-0.012	
BOOK				
Kindergarten	78.1	79.4	-1.3	7063
Spring 1st Grade	103.6	108.5	-4.9	7121
Spring 3rd Grade	117.8	129.2	-11.4†	7084
Spring 5th Grade	109.3	107.1	2.2	7124
PERMANENT CHARACTERISTICS				
ENTRY	65.7	65.6	0.1	7143
BIRTH WEIGHT	119.8	116.3	3.5‡	7045
AGE AT 1st BIRTH	23.9	24.0	-0.1	7059
WIC	0.431	0.422	0.09	7104
RACE[a]				
WHITE	0.649	0.617	0.032	
BLACK	0.154	0.162	-0.008	
HISPANIC	0.135	0.143	-0.008	
ASIAN	0.021	0.032	-0.011	

OTHER RACE	0.042	0.047	-0.005

Notes: MATH: standard normalized IRT math scores; SES: socioeconomic status; BOOK: the number of books that the child has; ENTRY: age (in months) at kindergarten entry; BIRTH WEIGHT: weight at birth (in ounces); AGE AT 1st BIRTH: the mother's age at first birth (in years); WIC: whether the child receives any WIC benefits as an infant or child. a: No standard errors are calculated. Panel weights are applied. ‡: p-value <0.01; †: p-value <0.05; *: p-value <0.10.

Table 3 The math gap at the mean in various specifications

Panel A

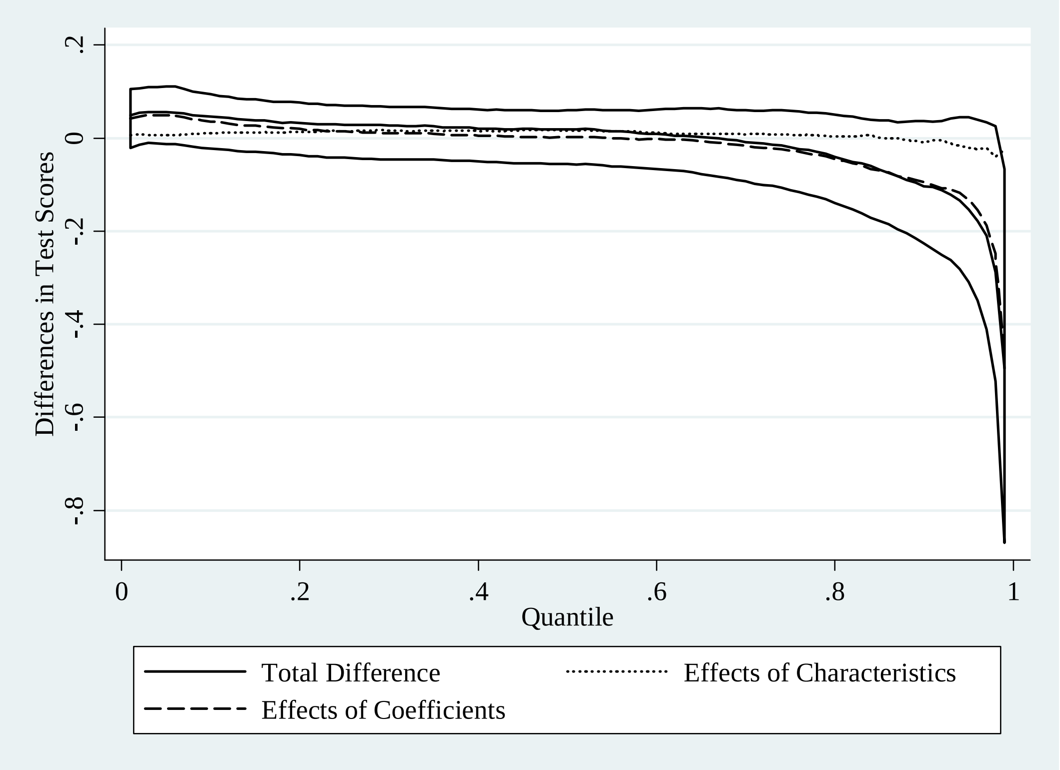
	Fall Kindergarten	Spring Kindergarten	Spring 1st Grade	Spring 3rd Grade	Spring 5th Grade
Raw Difference	-0.008 (0.036)	-0.046 (0.038)	-0.079 (0.038)†	-0.191 (0.040)‡	-0.195 (0.040)‡
N	7147	7147	7147	7147	7147
Between School Difference	-0.021 (0.030)	-0.053 (0.033)	-0.067 (0.034)†	-0.168 (0.035)‡	-0.185 (0.034)‡
Within School Difference	-0.061 (0.028)†	-0.069 (0.028)†	-0.090 (0.026)‡	-0.171 (0.028)‡	-0.166 (0.029)‡
N	6889	6889	6942	6905	6945

Panel B

	Fall Kindergarten		Spring Kindergarten		Spring 1st Grade		Spring 3rd Grade		Spring 5th Grade	
	Male[a]	Pooled[b]	Male	Pooled	Male	Pooled	Male	Pooled	Male	Pooled
Total	0.014 (0.036)		0.051 (0.038)		0.087 (0.038)†		0.191 (0.040)‡		0.201 (0.040)‡	
Characteristics[c]	-0.011 (0.020)	-0.006 (0.019)	-0.005 (0.020)	-0.001 (0.019)	0.015 (0.021)	0.021 (0.019)	0.021 (0.022)	0.025 (0.020)	0.011 (0.021)	0.018 (0.021)
Coefficients[d]	0.025 (0.031)	0.021 (0.030)	0.056 (0.033)*	0.052 (0.033)	0.072 (0.034)†	0.066 (0.033)†	0.170 (0.035)‡	0.166 (0.034)‡	0.190 (0.035)‡	0.183 (0.034)‡
Male N	3427		3427		3463		3442		3464	
Female N	3462		3462		3479		3463		3481	

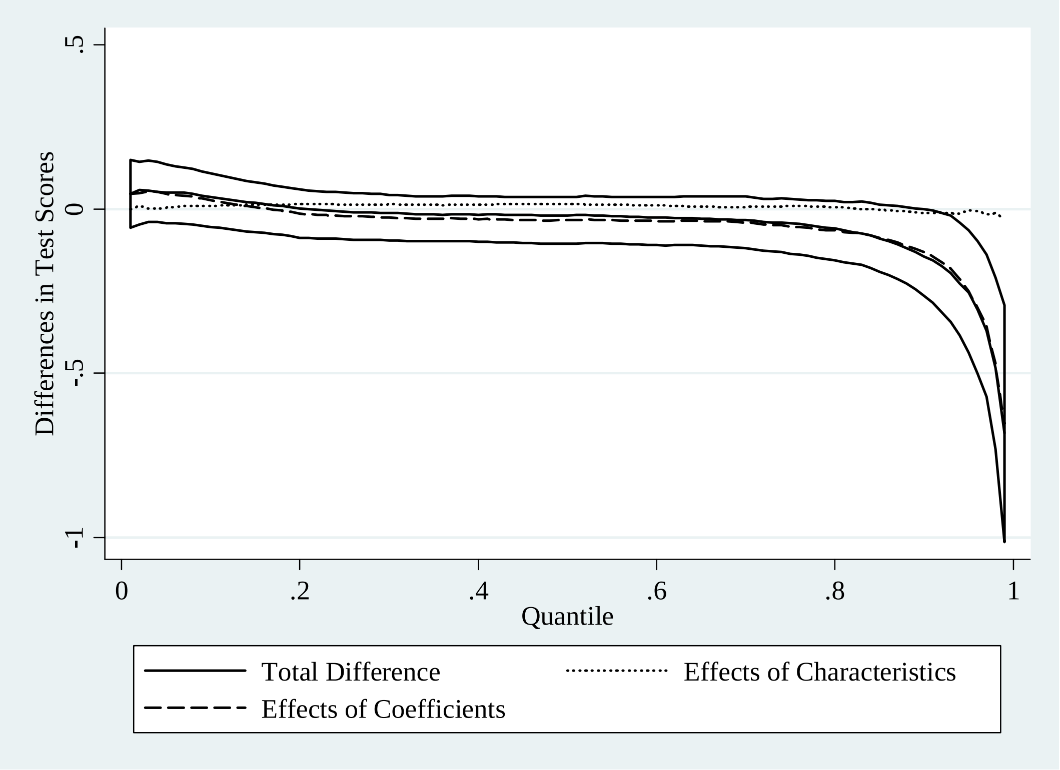
Notes: Raw Difference comes from Table 2. Between School Difference is estimated by (1). Within School Difference is estimated by (1) with school dummies added. a: The boys' coefficients are used as the norm in (2). b:The coefficients from the pooled sample are used as the norm in (2). c: The size of the total difference due to the difference in the independent variables. d: The size of the difference due to the difference in the coefficients. Panel weights are applied. ‡: p-value <0.01; †: p-value <0.05; *: p-value <0.10.

Figure 1 The gender math gap in the fall of kindergarten



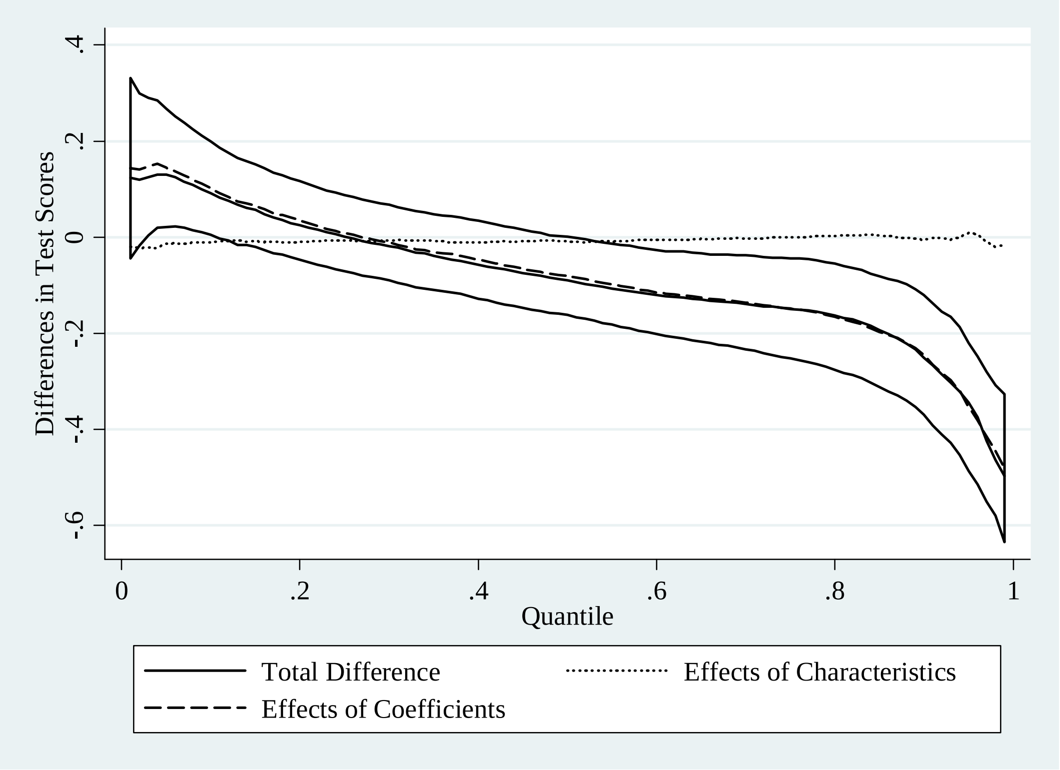
Notes: 95% Confidence interval for the effects of coefficients is estimated with 100 bootstrappings. Panel weights are applied.

Figure 2 The gender math gap in the spring of kindergarten



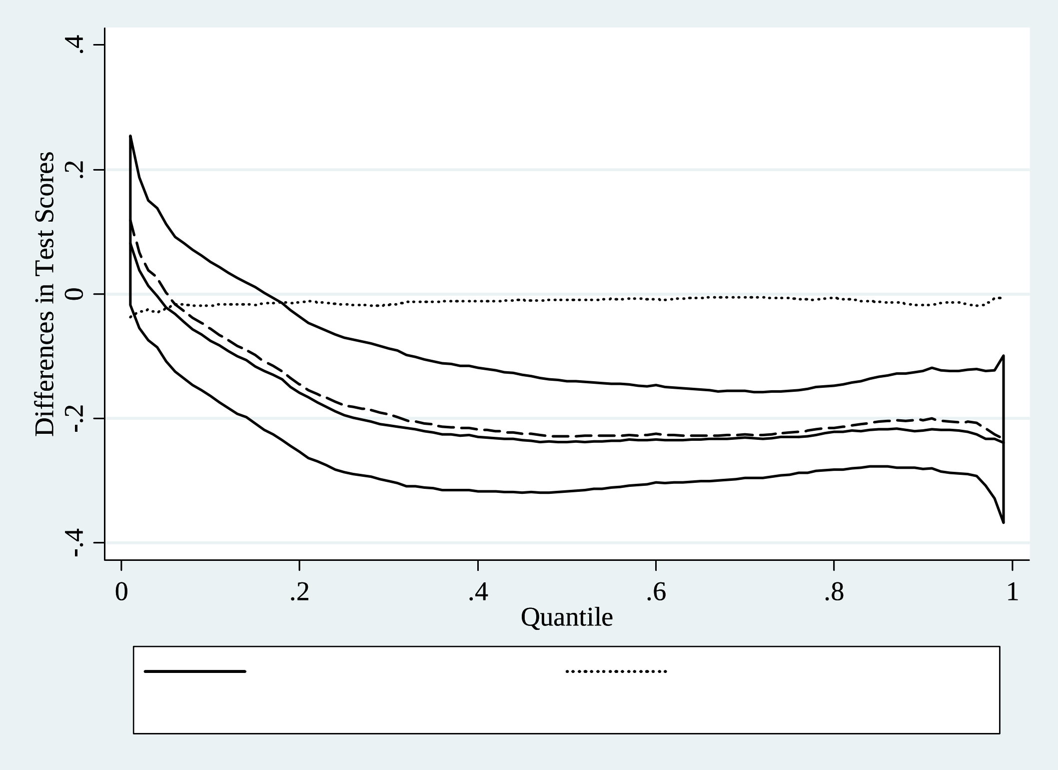
Notes: 95% Confidence interval for the effects of coefficients is estimated with 100 bootstrappings. Panel weights are applied.

Figure 3 The gender math gap in the spring of 1st grade



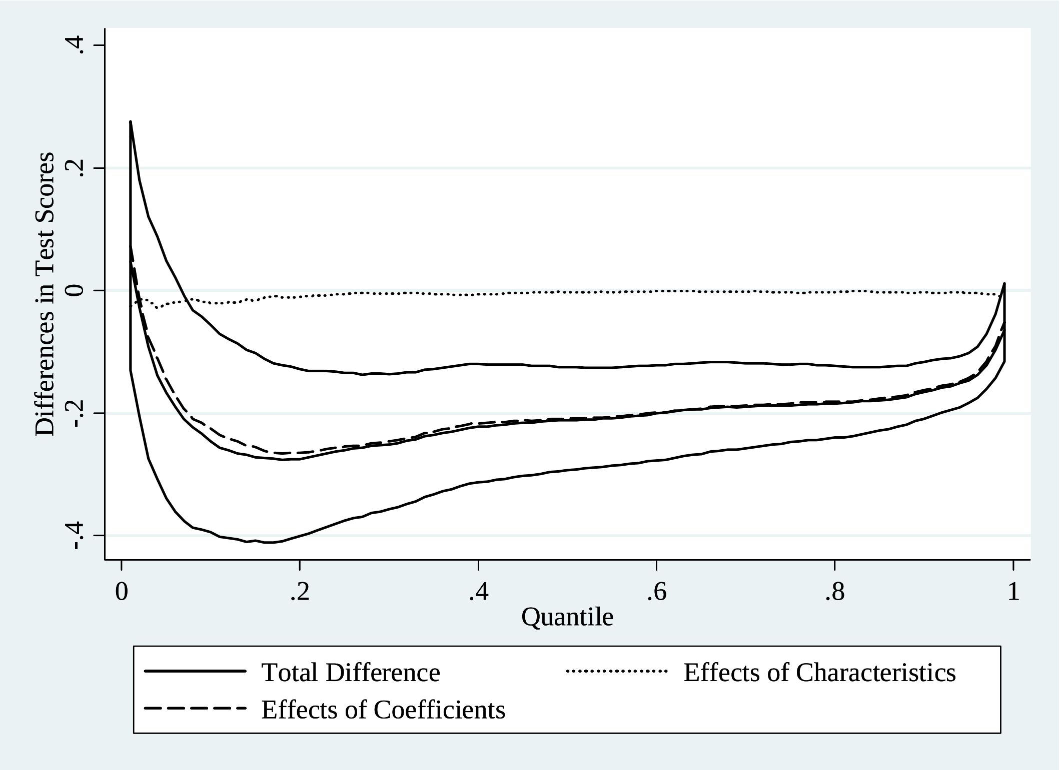
Notes: 95% Confidence interval for the effects of coefficients is estimated with 100 bootstrappings. Panel weights are applied.

Figure 4 The gender math gap in the spring of 3rd grade



Notes: 95% Confidence interval for the effects of coefficients is estimated with 100 bootstrappings. Panel weights are applied.

Figure 5 The gender math gap in the spring of 5th grade



Notes: 95% Confidence interval for the effects of coefficients is estimated with 100 bootstrappings. Panel weights are applied.