Discussion Paper Series – CRC TR 224

Discussion Paper No. 667
Project A 04

Double in Trouble: Boys and Learning in School in Texas, North Carolina, and Italy

Antonio Ciccone[1]
Federico Cingano[2]
Walter Garcia-Fontes[3]

March 2025

[1]University of Mannheim and CEPR, Email: antonio.ciccone@uni-mannheim.de
[2]Bank of Italy
[3]Universitat Pompeu Fabra and Barcelona School of Economics

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Double in Trouble: Boys and Learning in School in Texas, North Carolina, and Italy

Antonio Ciccone* Federico Cingano† Walter Garcia-Fontes‡

March 2025§

Abstract

Children have been found to learn less in school when there are more boys in the classroom. We make two contributions to the literature. First, we revisit the empirical approach based on within-school grade-level variation. We find that with grade retention (a feature of most school systems) the approach is subject to a selection bias that generates effects of boys on their classmates' skills in the absence of any peer effects. We propose an alternative based on variation across enrollment cohorts. Second, we implement both approaches using data on around 3.6 million children in 14,000 primary schools in Texas, North Carolina, and Italy. We find adverse effects of boys on how much children learn in school using both the grade-level and the enrollment-cohort approach. The enrollment-cohort approach yields adverse effects on boys that are at least double the adverse effects on girls in all three school systems. By contrast, consistent with the selection bias we highlight, the grade-level approach yields similar adverse effects on boys and girls in the school systems with substantial grade retention.

Keywords: Peer effects in school, grade retention, selection bias

JEL Codes: I21, J16, J24

---

*University of Mannheim and CEPR

†Bank of Italy

‡Universitat Pompeu Fabra and Barcelona School of Economics

§We thank Maximilian Bach, Gabrielle Fack, Libertad Gonzalez, Stephan Litschig, Barbara Petrongolo, and Alessandro Tarozzi for their comments. We also thank Alexander Göppert for excellent research assistance. This paper supersedes and replaces the 2015 mimeo "Peer Effects when Girls and Boys Learn Together in School, a Birth Cohort Approach" by Ciccone and Garcia-Fontes, which contained some of the theoretical analysis but did not have any empirical evidence for Texas, North Carolina, or Italy. We thank the staff at the University of Texas Dallas Education Research Center, the North Carolina Education Research Data Center, the Italian National Institute for the Evaluation of the Education System, and Ricardo Maertens for their help with the data. Financial support by the German Research Foundation (DFG) through CRC TR 224 (project A04) is gratefully acknowledged.

1 Introduction

Boys are born with greater risk of developing behavioral and socio-emotional problems than girls, and boys are also more adversely affected than girls by disadvantages in their family, social, and school environments (Tyre, 2008; Bertrand and Pan, 2013; Autor et al., 2016). As a result, boys are more likely than girls to generate disruption in the classroom (Lavy and Schlosser, 2011; Autor et al., 2019; Gong et al., 2021), and this leads to children learning less in school when there are more boys in the classroom (Hoxby, 2000, 2002; Lavy and Schlosser, 2011; Hu, 2015; Gong et al., 2021).

Whether more boys in the classroom affects how much boys and girls learn to the same extent is an open question. The literature has found the classroom learning environment to be worse when there are more boys (Lavy and Schlosser, 2011; Gong et al., 2021). A worse learning environment can be expected to have adverse effects on both boys and girls. However, the literature has also documented stronger negative effects of a disadvantaged school environment on boys than on girls (Diette and Oyelere, 2014; Autor et al., 2016). Existing estimates of the effect of boys on the skills of boys and girls in school depend on the empirical approach used. Studies relying on within-school random assignment across classrooms in China find 2 to 4 times larger adverse effects on the test scores of boys than girls (Hu, 2015; Gong et al., 2021). Previous work using within-school variation across cohorts in Texas and Israel found more similar effects on boys and girls (Hoxby, 2000, 2002; Lavy and Schlosser, 2011).[1]

We contribute evidence on the effect of boys on learning in school using data on around 3.6 million children in 14,000 primary schools in Texas, North Carolina, and Italy. While previous studies focused on one school system, we consider three school systems to check for a more general pattern. Our empirical approach is similar to Hoxby (2000) and Lavy and Schlosser (2011) in that we use within-school variation across cohorts. The main difference is that they examine how test scores depend on the variation in the share of boys and girls in the same grade. We propose an alternative intention-to-treat approach that examines how test scores depend on the variation in the share of boys and girls in the same enrollment cohort (as defined by the school system's primary-school enrollment policy[2]). We consider variation across enrollment cohorts because in school systems with grade retention (most school systems[3]) the grade-level approach of Hoxby (2000) and Lavy and Schlosser (2011) is subject to a simple selection bias. As a result of the selection bias, the grade-level approach may yield a positive effect of boys on the skills of boys in their class and a negative effect of boys on the skills of girls in their class even if there are no true peer effects of any type. Because of the selection bias, the grade-level approach could also indicate positive peer effects when true peer effects are negative and vice versa.

[1] While we focus on the effect of more boys, these studies look at the effect of more girls. As the shares of girls and boys in a cohort sum to one, the two approaches are equivalent (the effect of boys is always equal to minus the effect of girls).

[2] Using the enrollment policy of school systems as the basis of an intention-to-treat approach goes back to at least Bedard and Dhuey (2006). They examine the effect of maturity at the start of primary education on academic achievement.

[3] In the European Union, 24 of 27 countries have grade retention starting in primary school (European Commission, 2021). Outside the EU, almost all countries allow for grade retention (UNESCO, 2002).

Our results using the enrollment-cohort approach confirm the adverse effect of boys on how much children learn in school. This is the case in Texas, North Carolina, and Italy. We find a substantially stronger adverse effect of boys on boys than girls in all three school systems. The adverse effect of boys on the skills of boys is at least double the adverse effect on girls. This is consistent with the 2 to 4 times larger adverse effect on boys in studies relying on within-school random assignment across classrooms in Hu (2015) and Gong et al. (2021). When we instead use the grade-level approach, results for the three school systems depend on the extent to which there is grade retention. In Texas and North Carolina, the grade-level approach yields that boys have a similar adverse effect on the test scores of boys and girls. The difference from the results obtained using the enrollment-cohort is consistent with the selection bias of the grade-level approach we highlight when it is taken into account that grade retention in Texas and North Carolina is substantial. On the other hand, in Italy, the grade-level approach yields that the adverse effect of boys on boys relative to the adverse effect of boys on girls is similar to the enrollment-cohort approach. This is also consistent with the selection bias of the grade-level approach we highlight as there is relatively little grade retention in Italy.

In sum, we find adverse effects of boys on how much boys learn in school that are more than twice the adverse effects on girls in three different school systems. Our results are consistent with findings based on random assignment across classrooms (Hu, 2015; Gong et al., 2021) and findings that disadvantages in the family, social, and school environment have stronger adverse effects on boys than girls (Bertrand and Pan, 2013; Autor et al., 2016). While our results differ from Hoxby (2000) and Lavy and Schlosser (2011), we identify the methodological roots of this discrepancy. We show that in school systems with grade retention, Hoxby's and Lavy and Schlosser's grade-level approach understates the adverse effect of boys on boys and overstates the adverse effect on girls (Hoxby as well as Lavy and Schlosser study school systems with grade retention). Hence, our analysis reconciles existing results on the effects of boys on how much children learn in school and contributes new evidence for three different school systems.

We discuss the bias of the grade-level approach in school systems with grade retention in Section 3. The intuition can be illustrated in the following example. Consider a primary school enrolling exactly 20 children each year. Suppose there are no peer effects of any kind: the ability of children at enrollment is independent of all characteristics of their classmates and all children learn at the same pace in school. However, the ability distribution and the grade retention policy are such that, because of low skills, 10 percent of children are expected to be retained exactly once before they reach the test grade. Suppose also that the expected share of boys in each enrollment cohort is 50 percent. Hence, the unconditional expectation for the school is that 50 percent of children in the test grade are boys and that 10 percent of the boys and 10 percent of the girls in the test grade have been retained once in a lower grade.

Now suppose that an enrollment cohort with the expected number of 10 boys and 10 girls is followed by an enrollment cohort with 20 boys (for exogenous reasons). By the time the cohort that started with

20 boys reaches the test grade, the conditional expectation is that only 18 boys are left, as 2 have been retained because of their low skills. On the other hand, we expect that the cohort has seen the arrival of 1 retained boy and 1 retained girl from the enrollment cohort that started school one year earlier with 10 boys and 10 girls. Hence, once it has reached the test grade, we expect the cohort to consist of 18 non-retained boys, 1 retained boy, and 1 retained girl. Now suppose the skills of retained children are below those of non-retained children in the same grade (the case in our data).[4] The relatively low share of retained boys among boys (1/19) and the relatively high share of retained girls among girls (1/1)—when compared to the unconditional expectation (1/10)—imply that we expect boys to perform better on the test than the unconditional expectation for the school, and girls to perform worse. As this coincides with a high share of boys in the test grade, the grade-level approach will lead to the conclusion that there is a positive effect of boys on the skills of boys and a negative effect on the skills of girls.

The next section reviews the literature. Section 3 discusses the grade-level and enrollment-cohort approach to peer effects within schools in a model with grade retention. We also discuss an approach where the grade-level share of boys is instrumented by the share in the enrollment cohort. Sections 4 and 5 describe our data and summarize our empirical results. The last section concludes.

2 Related Literature

We contribute to the literature on peer effects, see Sacerdote (2014), Epple and Romano (2011) and Cools and Patacchini (2021) for reviews. We are most closely related to the work on gender peer effects of Hoxby (2000), Whitmore (2005), Lavy and Schlosser (2011), Hu (2015), and Gong et al. (2021). Overall, these studies find an adverse effect of boys on learning in school.[5] This appears to be because of a worse classroom learning environment (Lavy and Schlosser, 2011; Gong et al., 2021).

Where there is some disagreement is in whether boys and girls are affected to the same extent by boys in the classroom. Hoxby (2000) and Lavy and Schlosser (2011) find adverse effects of boys on the skills of boys and girls that are of a similar magnitude in Texas and Israel respectively. Their estimation approach uses within-school variation in the share of boys and girls at a given grade level. By contrast, the results of Hu (2015) and Gong et al. (2021) indicate significantly stronger adverse effects of boys on the skills of boys than girls. Gong et al. (2021) find that the effect of a higher share of male students on the test scores of male students is more than twice the effect on female students, while Hu (2015) finds a ratio of around 4. Their estimates rely on within-school random assignment across classrooms in

[4] We find that retained children do worse on average in standardized tests than their non-retained peers in the same grade, although retained children will generally have had one more year of schooling.

[5] Comparing the size of estimates of these studies is difficult as different studies consider different grade levels (from the start of primary education to high school) and use country-specific standardized tests. Our comparisons across school systems therefore focus on the difference in the effect of boys on boys and on girls.

China.[6] Our findings point in the same direction.[7]

A related body of literature uses within-school variation in peer gender across cohorts to investigate whether gender peer effects influence subsequent educational choices. Recent studies focus on the likelihood of girls choosing male-dominated STEM fields, which are typically associated with higher wages. Schneeweis and Zweimüller (2012) and Schøne et al. (2019) report significant positive effects of female peers on girls' enrollment in male-dominated (e.g., vocational) schools in Austria and Norway. In the case of university choices, Brenøe and Zølitz (2022) find that female peer presence in high schools in Denmark negatively affects girls' enrollment and graduation rates in STEM fields, while the effect on boys is positive. Anelli and Peri (2019) report no significant gender peer effects on college choices among Italian students of either gender. At the PhD level, Bostwick and Weinberg (2022) find that female peers positively influence girls' probability of on-time program completion.

Further research utilizes detailed administrative linked data to explore gender peer effects in labor market outcomes and fertility decisions. Looking at within-school variation in Danish schools, Brenøe and Zølitz (2022) find that women who had more female peers are less likely to work in STEM occupations, earn less by age 36, and have more children. In Norway, Black et al. (2013) find a positive effect of female peers on women's full-time participation and earnings but no significant effect on men.

3 Estimating Peer Effects with Grade Retention

We now consider three different approaches to estimate gender peer effects based on within-school variation in the share of boys and girls. First, the grade-level approach where children's skills are related to the share of boys in their grade using a least-squares approach. Second, an instrumental-variables approach where the grade-level share of boys is instrumented using the share of boys in the corresponding enrollment cohort. Third, the enrollment-cohort approach, an intention-to-treat approach where children's skills are related to the share of boys in their enrollment cohort. In school systems without grade retention, the three approaches will yield identical and consistent estimates of peer effects. Differences between the three approaches emerge when they are applied to school systems with grade retention.

3.1 Primary School with Grade Retention

The simplest model that illustrates the issues at the core of our analysis has one primary school and two grade levels, low grade (LG) and high grade (HG). Each year a continuum of children enrolls in school.

[6] Whitmore (2005) uses random variation in the share of girls and boys in the classroom generated by Project STAR. When she examines effects on boys and girls separately, her estimates become noisy and mostly statistically insignificant (sample size is small compared to other studies). Less closely related are Lu and Anderson (2015) and Hill (2015). They examine the effect of male versus female seating arrangements within classrooms and male versus female friends respectively.

[7] Bach and Sievert (2024) adapt our theoretical model to discuss a similar bias that arises when the grade-level approach is used to estimate class size effects in school systems with grade retention.

Enrollment cohort t starts school in year t. Children differ in their skills when they enroll in school. We assume the same skill distribution across enrollment cohorts for girls and boys. Enrollment cohorts solely differ in the share of girls and boys. We take this variation to reflect exogenous factors.

All children accumulate the same amount of skills in LG. At the end of their first year in LG, children i are promoted to HG if their skills a_i are above a threshold p

$$a_i \geq p. \tag{1}$$

The threshold p is the same for girls and boys. Children with skills below the threshold p at the end of their first year in LG are retained for one more year in LG before being promoted to HG. Our assumptions imply that the same share of girls and boys will be retained. We will denote this share by ρ. Retained children accumulate skills λ in their second year in LG. We assume that children are never retained in HG. At the end of HG, children take a test measuring their skills.

The gender peer effect in the model arises because learning in HG may depend on the share of girls and boys in the grade. We assume that HG skills accumulated by boys are $\alpha + \beta_{boy} \cdot ShareBoys_\tau$, where $ShareBoys_\tau$ is the share of boys in HG in the school year starting in calendar year τ. Analogously, girls accumulate skills $\alpha + \beta_{girl} \cdot ShareBoys_\tau$. Hence, the gender peer effect may affect the learning of girls and boys differently.

The expected skills, and hence expected test performance, of children at the end of HG will differ for boys and girls because they may be affected differently by the gender peer effect. Expected skills and test performance at the end of HG will also differ depending on whether children were retained or not in LG. For example, the expected skills and test performance of boys who have not been retained previously is

$$E\big(test_{boy,\tau}^{nonret} \big| ShareBoys_\tau\big) = E(a\,|a \geq p) + \alpha + \beta_{boy} \cdot ShareBoys_\tau. \tag{2}$$

The term $E(a\,|a \geq p)$ captures that grade retention induces positive selection of the children who reach HG without being retained in LG. The last term captures the gender peer effect in HG. The expected test performance of boys in HG who have been retained previously is

$$E\big(test_{boy,\tau}^{ret} \big| ShareBoys_\tau\big) = E(a\,|a < p) + \lambda + \alpha + \beta_{boy} \cdot ShareBoys_\tau. \tag{3}$$

The term $E(a\,|a < p)$ captures that grade retention induces negative selection of the children in HG who were retained in LG. λ denotes the skills accumulated by retained children in their second year in LG. The last term captures the gender peer effect in HG. The expected test performance of girls can be obtained analogously and only differs in the parameter capturing the gender peer effect.

Children who have been retained in a previous grade perform worse in standardized tests than non-retained children in our data. To ensure that this is also the case in our model, we assume

$$\begin{aligned}\Delta_{ret}^{nonret} &= E\big(test_{boy,\tau}^{nonret}\big|ShareBoys_\tau\big) - E\big(test_{boy,\tau}^{ret}\big|ShareBoys_\tau\big) \\ &= E(a\,|a \geq p) - E(a\,|a < p) - \lambda > 0.\end{aligned} \tag{4}$$

where we use (2) and (3). As the threshold p and the distribution of skills a are the same across enrollment cohorts and for girls and boys, (4) implies that the skill gap between non-retained and retained children in HG, and hence the gap in test performance, does not vary over time and is the same for boys and girls.

The variation in the share of boys in HG $ShareBoys_\tau$ across school years τ is driven by the exogenous variation in the share of boys across enrollment cohorts

$$ShareBoys_\tau = (1-\rho)\cdot\phi^{\tau-1} + \rho\cdot\phi^{\tau-2} \tag{5}$$

where ρ is the share of children retained after their first year in LG and $\phi^{\tau-1}$ and $\phi^{\tau-2}$ denote the share of boys in the cohort that enroll in primary school in calendar year $\tau-1$ and $\tau-2$ respectively.

This completes the description of the model. The exogenous gender composition of enrollment cohorts determines the gender composition in HG and the skills, and hence test performance, of all children in HG.

What we want to highlight with this model is that grade retention implies that children's skills in HG will covary systematically with the gender composition in HG absent any peer effects. To see this, it is useful to write the expected test performance of boys as a weighted average of the test performance of non-retained boys in (2) and retained boys in (3) with weights equal to the shares of non-retained and retained boys respectively. This simplifies to

$$\begin{aligned}E\big(test_{boy,\tau}\big|ShareBoys_\tau, NonRetBoys_\tau\big) = E(a\,|a < p) + \alpha + \lambda + \beta_{boy}\cdot ShareBoys_\tau \\ + NonRetBoys_\tau\cdot\Delta_{ret}^{nonret}\end{aligned} \tag{6}$$

where $\Delta_{ret}^{nonret} > 0$ is defined in (4), and

$$NonRetBoys_\tau = \frac{\textit{Non-Retained Boys}_\tau}{\textit{All Boys}_\tau} = \frac{(1-\rho)\cdot\phi^{\tau-1}}{(1-\rho)\cdot\phi^{\tau-1} + \rho\cdot\phi^{\tau-2}}. \tag{7}$$

Now the effect we want to highlight is immediate. According to (7), a higher share of boys in enrollment cohort $\tau-1$ results in a higher share of non-retained boys in HG in school year τ. According to (6), this translates into a higher expected test performance of boys in HG even in the absence of any peer effects as $\Delta_{ret}^{nonret} > 0$. In (5), we already saw that a higher share of boys in enrollment cohort $\tau-1$ also increases the share of boys among all children in HG in school year τ. Hence, both the expected test performance of boys and the share of boys among children in HG in school year τ increase in response to a higher share of boys in enrollment cohort $\tau - 1$ even if there are no peer effects. While there is a countervailing effect as (6) and (7) also imply that an increase in the share of boys in enrollment cohort $\tau - 2$ increases the share of retained boys in HG in school year τ and therefore lowers the expected test performance of boys, we show below that this effect is dominated as long as the share of children retained in LG is below one half. As a result, using the grade-level approach will result in a spurious positive peer effect of boys on boys as long as Δ_{ret}^{nonret} is positive and ρ smaller than one half. For girls, the same argument yields that the expected test performance of girls in HG and the share of girls among children in HG also move in the same direction. As the share of girls covaries negatively with the share of boys, this implies that the expected test performance of girls in HG and the share of boys among children in HG move in opposite directions. When using the grade-level approach this will result in a spurious negative peer effect of boys on girls.

3.2 Estimation at the Grade Level

We now examine different within-school estimation approaches. To do so, we take the exogenous share of boys in each enrollment cohort to fluctuate randomly around 1/2, $\phi^t = 1/2 + \epsilon^t$ where ϵ is i.i.d. with $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2$. Combining (5), (6), (7) linearized around $\phi = 1/2$, and the standard LS formula yields the following LS slope when the test performance $test_{boy,\tau}^i$ of boys i in HG in school year τ is regressed on the share of boys in HG

$$\hat{b}_{boy}^{grade} = \frac{Cov(test_{boy,\tau}^i, ShareBoys_\tau)}{Var(ShareBoys_\tau)} = \beta_{boy} + \left(\frac{1}{2} - \rho\right)\frac{4\rho(1-\rho)}{\rho^2 + (1-\rho)^2} \cdot \Delta_{ret}^{nonret} \tag{8}$$

where $\Delta_{ret}^{nonret} > 0$ is defined in (4).[8] Hence, the grade-level approach yields an upward biased estimate of the peer effect of boys on boys β_{boy} when there is grade retention, $\rho > 0$, as long as less than half the children are retained before they reach HG ($\rho < 1/2$). The result in (8) is mostly as expected given what we highlighted above: variation in the share of boys in enrollment cohort $\tau - 1$ induces a positive covariation between the share of boys in HG and their expected skills even absent any peer effects if $\Delta_{ret}^{nonret} > 0$. However, for the bias in (8) to be upward, it must also be the case that the share of children retained before they reach HG is smaller than one half. This condition reflects a countervailing effect to the positive covariation between the share of boys in HG and their expected skills. The countervailing effect arises because variation in the share of boys in enrollment cohort $\tau - 2$

[8] The simplest derivation of (8) uses that the test scores of boys can be written as $test_{boy,\tau}^i = E\big(test_{boy,\tau}^i \,|\, \phi^{\tau-1}, \phi^{\tau-2}\big) + \eta_\tau^i$ where by the law of iterated expectations $E\big(\eta_\tau^i \,|\, \phi^{\tau-1}\big) = E\big(\eta_\tau^i \,|\, \phi^{\tau-2}\big) = 0$. Moreover, (5) and (6) combined with the linearized version of (7) imply that $E\big(test_{boy,\tau}^i \,|\, \phi^{\tau-1}, \phi^{\tau-2}\big) = E(a\,|\,a < p) + \lambda + \alpha + \beta_{boy} \cdot \big((1-\rho)\cdot\phi^{\tau-1} + \rho\cdot\phi^{\tau-2}\big) + 1/2 \cdot (1-\rho)\big((1+\rho)\cdot(\phi^{\tau-1} - 1/2) - \rho\cdot(\phi^{\tau-2} - 1/2)\big)\cdot\Delta_{ret}^{nonret}$. Combining this last expression with (5) yields (8).

induces a negative covariation between the share of boys in HG and their expected skills. When $\rho < 1/2$, the positive covariation dominates and the grade-level approach yields an upward biased estimate of the peer effect of boys on boys. In our data, the share of children retained before they reach the test grade is much smaller than one half. We therefore assume $\rho < 1/2$.

The LS slope when the test performance $test^i_{girls,\tau}$ of girls i in HG in school year τ is regressed on the share of boys in HG can be derived analogously as

$$\hat{b}^{grade}_{girl} = \frac{Cov(test^i_{girls,\tau}, ShareBoys_\tau)}{Var(ShareBoys_\tau)} = \beta_{girl} - \left(\frac{1}{2} - \rho\right)\frac{4\rho(1-\rho)}{\rho^2 + (1-\rho)^2} \cdot \Delta^{nonret}_{ret}. \tag{9}$$

Hence, the grade-level approach yields a downward biased estimate of the peer effect of boys on girls β_{girls} if there is grade retention as long as less than half of the children in an enrollment cohort are retained (and $\Delta^{nonret}_{ret} > 0$). The biases when using the grade-level approach to estimate the gender-peer effect of boys on girls in (9) and the gender-peer effect of boys on boys in (8) are of opposite sign but of the same magnitude.[9]

In sum, the grade-level approach overstates the peer effect of boys on boys and understates the peer effect of boys on girls when there is grade retention and retained students have lower skills than non-retained students in the same grade. One might think that the difference between the skills of non-retained and retained students could be accounted for in the empirical analysis by controlling for whether students were retained or not. However, while the literature using the grade-level approach controls for the effect of predetermined individual characteristics on individual test performance, we are not aware of empirical work controlling also for whether a child has been retained. The reason is probably that, generally, whether a child is retained or not is endogenous.

3.3 An Instrumental-Variables Approach at the Grade Level

The bias of the grade-level approach is not eliminated using an instrumental-variables approach where the share of boys in HG is instrumented by the (exogenous) share of boys in the corresponding enrollment cohort (the enrollment cohort of non-retained children in HG). This is unsurprising by now, as we have shown that the share of boys in the enrollment cohort is correlated with the share of non-retained boys and non-retained girls in HG one year later and that the instrument therefore does not satisfy the exclusion restriction.[10] Maybe more surprisingly, the IV-bias may be even larger than the bias of the

[9] The bias for the gender-peer effect of boys on girls and the gender-peer effect of boys on boys is of opposite sign but of the same magnitude because we assumed the same retention rate for boys and girls and the same test-score gap between non-retained and retained students. Without these assumptions, the magnitude of the bias is the gender-specific test-score gap multiplied by $4\rho_x(1-\rho_x)(1/2-\bar{\rho})/(\bar{\rho}^2 + (1-\bar{\rho})^2)$ where $\bar{\rho}$ is the average and ρ_x the gender-specific retention rate. The sign of the bias remains unchanged.

[10] This follows from (6) and (7), as the share of boys in enrollment cohorts affects expected test performance in HG through both the share of boys in HG and the share of non-retained boys in HG. Hence, the exclusion restriction—that the variation in the share of boys across enrollment cohorts affects third-grade test scores solely through the share of boys in third grade—is not satisfied.

grade-level approach. For example, using the standard formula for IV estimates and focusing on the result for boys, the bias of the IV approach is

$$BiasIV_{boy} = \frac{Cov(test^{i}_{boy,\tau}, \phi^{\tau-1})}{Cov(ShareBoys_{\tau}, \phi^{\tau-1})} - \beta_{boy} = 2\rho \cdot \Delta^{nonret}_{ret}. \tag{10}$$

Hence, the IV approach also yields a positive peer effect of boys on boys in the absence of any true peer effects and the direction of the bias is the same as when using the grade-level approach.

To get an idea of the size of the IV bias in (10) as compared to the bias of the grade-level approach in (9), we draw on our data for children in third grade in Texas primary schools. Around 15% of third graders have been retained in a lower grade, and non-retained children in third grade perform around 0.3-0.4 standard deviations better than retained children in the standardized test. As a result, the implied size of the bias in (9) is around 0.07-0.1. The IV bias implied by (10) is 0.1-0.14 and therefore larger.

3.4 Estimation at the Enrollment-Cohort Level

As both the grade-level approach in Section 3.2 and the IV approach in Section 3.3 may yield spurious gender peer effects, we consider an alternative that we call the enrollment-cohort approach to gender peer effects. This approach involves regressing the test performance of children in HG on the share of boys in their enrollment cohort. As non-retained children in enrollment cohort t will start HG in year $t+1$ and retained children in year $t+2$, the expected test performance in HG of boys in enrollment t will be $Etest^{t}_{boy} = (1-\rho)E\big(test^{nonret}_{boy,t+1}\big|ShareBoys_{t+1}\big) + \rho E\big(test^{ret}_{boy,t+2}\big|ShareBoys_{t+2}\big)$ where ρ is the share of retained boys. Substituting (2) and (3) yields the expected test performance of boys in enrollment cohort t

$$Etest^{t}_{boy} = Ea + \alpha + \rho\lambda + (1-\rho)\beta_{boy} \cdot ShareBoys_{t+1} + \rho\beta_{boy} \cdot ShareBoys_{t+2} \tag{11}$$

as non-retained boys in enrollment cohort t start HG in year $t+1$, while retained boys start in year $t+2$. Combining (5) and (11) yields that the LS slope when regressing the test performance $test^{it}_{boy}$ of boys i in enrollment cohort t on the share of boys in the enrollment cohort is[11]

$$\hat{b}^{enroll}_{boy} = \frac{Cov(test^{it}_{boy}, \phi^{t})}{Var(\phi^{t})} = \beta_{boy}\big(\rho^2 + (1-\rho)^2\big). \tag{12}$$

The LS slope when regressing the test performance of girls in enrollment cohort t on the share of boys in the enrollment cohort can be obtained analogously

$$\hat{b}^{enroll}_{girl} = \frac{Cov(test^{it}_{girl}, \phi^{t})}{Var(\phi^{t})} = \beta_{girl}\big(\rho^2 + (1-\rho)^2\big). \tag{13}$$

[11] The derivation is analogous to the grade-level approach.

Hence, just as for the grade-level approach, estimates obtained using the enrollment-cohort approach do not have a structural interpretation. However, according to (12) and (13), the enrollment-cohort approach has some important advantages when compared to the grade-level approach. First, the enrollment-cohort approach will not yield gender peer effects if there are no true gender peer effects, i.e. if $\beta_{boy} = \beta_{girl} = 0$. Second, the enrollment-cohort approach will yield positive (negative) gender peer effects if true gender peer effects are positive (negative). Third, the enrollment-cohort approach will yield a ratio of the gender peer effect of boys on boys and the gender peer effect of boys on girls equal to the true ratio of gender peer effects.[12]

4 Data Sources and Description

We use (confidential) data for three different school systems: Texas, North Carolina, and Italy.[13]

The data for Texas is provided by the Texas School Project at the University of Texas at Dallas and refers to all public and chartered primary schools. The assessment data we use is the Texas Assessment of Knowledge and Skills (TAKS). We use the TAKS standardized test scores in mathematics and reading administered in third grade. We match the TAKS assessment data with the student's month and year of birth, race, gender, and eligibility for a subsidized lunch from the Texas Education Agency's Public Education Information Management System. We implement the grade-level approach using the data for the four school years starting between 2005 and 2008. Enrollment cohorts are defined by the primary-school enrollment policy in Texas. The policy is that to be eligible to enroll in primary school, children must be at least five years old on September 1 of the school year. We therefore assign children born between September 1 of one year and August 31 of the next year to the same enrollment cohort. We do so starting from the set of students who attended third grade in a given school between the 2003/2004 and 2011/2012 school years. We then construct four enrollment cohorts, defined as children who should have entered third grade in the school years starting in 2005, 2006, 2007, or 2008, assuming they began primary school according to the enrollment rule and were neither retained nor skipped a grade.[14] In our data, approximately 85 percent of students who should be in third grade—had they followed the enrollment rule and not been retained or skipped a grade—are indeed in third grade. The remaining

[12] This last result only holds when retention rates are the same for boys and girls. However, for the gender gap in retention rates in our data for Texas and North Carolina (the retention rate of girls is some 5 percentage points lower than the retention rate of boys), the result continues to hold to a close approximation. To see this, note that with gender-specific retention rates, the enrollment-cohort approach yields the following estimates $\beta_{boy}\big(\rho_{boy}^2 + (1-\rho_{boy})^2\big)$ for the gender-peer effect of boys on boys and $\beta_{girl}\big(\rho_{girl}\rho_{boy} + (1-\rho_{girl})(1-\rho_{boy})\big)$ for the gender-peer effect of boys on girls. In our data for Texas, the retention rate for boys is 18 percent and the retention rate for girls is 13 percent. Substituting yields that the enrollment-cohort approach yields a ratio of the gender peer effect of boys on boys and of boys on girls equals to 0.96 times the true ratio.

[13] Hoxby (2000) also estimates gender peer effects in Texas primary schools, but her data refers to the 1990s when a different assessment system was used (the Texas Assessment of Academic Skills).

[14] We end up with four enrollment cohorts although we have the TAKS third-grade data for the school years between 2003 and 2011 because we want to ensure that we catch children who enroll early and/or skip a grade and children who enroll late and/or are retained.

15 percent would be in fourth or fifth grade if they had followed the enrollment rule and had not been retained in primary school.[15,16]

For North Carolina, the data is provided by the North Carolina Education Research Data Center at Duke University and, just as for Texas, refers to all public and chartered primary schools. The assessment data comes from the North Carolina End of Grade Test. We use the standardized test scores in mathematics and reading administered in third grade. The data also includes the student's month and year of birth, race, and gender. We implement the grade-level approach using the data for the 10 school years starting between 1999 and 2008. Enrollment cohorts are defined by the primary-school enrollment policy in North Carolina. The policy in place for the children in our data is that children must be five years old on October 15 of a school year to be eligible for admission in a school district. As we only have children's month of birth, we drop children born in October from the dataset and assign children born between November 1 of one year and September 30 of the next year to the same enrollment cohort.[17] We obtain enrollment cohorts using the same approach implemented for Texas. We start from the set of students who attended third grade in a given school between the 1997/1998 and 2011/2012 school years. We then construct 10 enrollment cohorts, defined as children who should have entered third grade between 1999 and 2008, assuming they began primary school according to the enrollment rule and were neither retained nor skipped a grade. In our data, approximately 83 percent of students who should be in third grade—had they followed the enrollment rule and not been retained or skipped a grade—are indeed in third grade. The remaining 17 percent would be in fourth or fifth grade if they had followed the enrollment rule and had not been retained in primary school.[18,19]

For Italy, the data is provided by the Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI). We use the standardized test scores in mathematics and reading administered in second grade (there is no test administered in third grade). The data also provides the student's month and year of birth, gender, and the educational attainment of mother and father (primary education, lower secondary education, upper secondary education, and college). We implement the grade-level approach using the data for the five school years starting between 2012 and 2016. Enrollment cohorts are defined by the primary-school enrollment rule in Italy. The rule in place for the children in our data is that children must be six years old on December 31 of a school year

[15] The data does not allow us to determine whether these students started primary school late or were retained in primary school. There are also some students—less than 0.5 percent—who if they had followed the enrollment rule and had not skipped a grade in primary school, would be in second grade. The data does not allow us to determine whether these students started primary school early or skipped a grade in primary school.

[16] Students who were retained in third grade—less than 1 percent—take the third-grade test twice. We use their score from the second time they take the test. Using their score from the first time yields nearly identical results.

[17] In 2007, North Carolina passed legislation that moved the date to August 31 starting in 2009.

[18] As with Texas, the data for North Carolina does not allow us to determine whether these students started primary school late or were retained. Additionally, some students—less than 0.5 percent—would be in second grade had they followed the enrollment rule and not skipped a grade in primary school.

[19] Students who were retained in third grade—less than 1 percent—take the third-grade test twice. We use their score from the second time they take the test. Using their score from the first time yields nearly identical results.

for admission to primary school. However, children who turn six before April 30 are allowed to enroll one school year earlier. We therefore drop children born January-April from the dataset and assign children born between May 1 and December 31 to the same enrollment cohort. We obtain enrollment cohorts using the same approach implemented for Texas and North Carolina. Starting with students who attended second grade in a given school between the 2011/2012 and 2018/2019 school years, we construct five enrollment cohorts. These cohorts consist of children who, according to the enrollment policy, should have started second grade between 2012 and 2016, assuming they were neither retained nor skipped a grade. In our data, approximately 98 percent of students who should be in second grade—had they followed the enrollment rule and not been retained or skipped a grade—are indeed in second grade. The remaining 2 percent would be in fourth or fifth grade if they had followed the enrollment rule and had not been retained in primary school.[20]

5 Estimating Equations and Empirical Results

We estimate the effect of boys on learning in primary school both using the enrollment-cohort approach and the grade-level approach. The estimating equation for the enrollment-cohort approach is

$$test_{gs}^{it} = \alpha_{gs} + \alpha_{gt} + \alpha_g X_{gs}^{it} + \beta_g EnrollmentCohortBoyShare^t + v_{gs}^{it} \quad (14)$$

where $test_{gs}^{it}$ refers to individual standardized test scores in third grade, i to individuals, t to enrollment cohorts, g to gender, and s to schools; X_i is a vector collecting predetermined characteristics of student i; and v_{gs}^{it} is the residual. The estimating equation for the grade-level approach is

$$test_{gs\tau}^{i} = \alpha_{gs} + \alpha_{g\tau} + \alpha_g X_{gs\tau}^{i} + \beta_g GradeLevelBoyShare_\tau + v_{gs\tau}^{i} \quad (15)$$

where $test_{gs\tau}^{i}$ refers to individual standardized test scores in third grade, i to individuals, τ to school years, g to gender, and s to schools; X_i is a vector collecting predetermined characteristics of student i; and $v_{gs\tau}^{i}$ is the residual.

Tables 1-3 contain our estimates of the effect of boys on the test scores of boys and girls in Texas, North Carolina, and Italy.[21] Panels A contain our estimates for the enrollment-cohort approach and Panels B for the grade-level approach. All tables report point estimates for the effect of the share of

[20] Approximately 0.5 percent of students would be in second grade if they had followed the enrollment rule and had not skipped a grade in primary school.

[21] In the Appendix Table we show that predetermined skills are balanced at the level of enrollment cohorts. We do so in two steps. We first obtain the predicted third-grade test scores of all children based on their predetermined characteristics. This is our measure of children's predetermined skills. Then we examine the effect of the share of boys in the enrollment cohort on the predetermined skills of boys and girls in the enrollment cohort. Although our estimates are quite precise, we find statistically insignificant effects of the share of boys in enrollment cohorts on the predicted test scores of boys and girls in the enrollment cohort, indicating that predetermined skills are balanced at the enrollment-cohort level.

boys on the test performance of boys and girls. Standard errors are clustered at the school level.

Table 1 reports our findings for Texas. Panel A contains our results using the enrollment-cohort approach. Estimates are based on almost 1.2 million children in more than 4,300 schools. The effect of the share of boys in the enrollment cohort on the test scores of girls in columns (1)-(4) is around -0.05 to -0.06. This implies that a 10 percentage points higher share of boys lowers the test scores of girls by around 0.5 percent of a standard deviation. Although this effect is small, it is statistically significant at the 10-percent level for reading. For mathematics, the effect is borderline statistically insignificant. The effect of the share of boys in the enrollment cohort on the test score of boys in columns (5)-(8) is -0.12 to -0.13. This implies that a 10 percentage points higher share of boys lowers the test scores of boys by around 1.2 percent of a standard deviation. The effects are statistically significant at the 1-percent level for mathematics and for reading. Comparing the same specifications with controls for predetermined characteristics yields adverse effects of boys on boys that are around 2.5 times the adverse effects of boys on girls. This difference is statistically significant at the 10-percent level.

Panel B of Table 1 reports our estimates for Texas using the grade-level approach. The effect of the share of boys in the grade on the test scores of girls in columns (1)-(4) is around -0.08 to -0.1. The effect of boys on the test score of boys in columns (5)-(8) is -0.05 to -0.08. Hence, the grade-level approach yields similar effects of more boys in the grade on the test scores of girls and boys (the difference is not statistically significant). The size of the effect is quite small, a 10 percentage points higher share of boys lowers the test scores of girls by around 0.5-0.8 percent of a standard deviation.[22] Notice that the main difference between the grade-level estimates in Panel B and the enrollment-cohort estimates in Panel A is that the enrollment-cohort approach yields a larger adverse effect of boys on boys and a smaller adverse effect of boys on girls. This is consistent with the selection bias of the grade-level approach in school systems with substantial retention (around 15% of the children in third grade in Texas have been retained in a lower grade).

Table 2 reports our findings for North Carolina. Panel A of Table 2 contains our results using the enrollment-cohort approach. Estimates are based on almost 900,000 children in more than 1,500 schools. The effect of the share of boys in the enrollment cohort on the test scores of girls in columns (1)-(4) is around -0.03 to -0.06. This effect is small and similar to what we obtained for Texas (but now the effect for mathematics is statistically significant at the 10-percent level and the effect for reading insignificant). The effect of boys on the test score of boys in columns (5)-(8) is -0.12 to -0.15, somewhat stronger than what we obtained for Texas. The effects are statistically significant at the 1-percent level for mathematics and for reading. Comparing the same specifications with controls for predetermined characteristics yields adverse effects of boys on boys that are around 2-3 times the adverse effects of boys

[22]These effects are less than half of what Hoxby (2000) estimates for Texas primary schools in the 1990s. However, she does not have individual student data and therefore estimates effects using average test scores in third grade.

on girls. This difference is statistically significant at the 10-percent level.

Panel B of Table 2 reports results for North Carolina using the grade-level approach. The effect of the share of boys in the grade on girls in columns (1)-(4) is around -0.04 to -0.1. The effect on boys in columns (5)-(8) is -0.03 to -0.08. Hence, the grade-level approach yields similar effects of boys on the test scores of girls and boys (the difference is not statistically significant). The size of the effects is similar to what we obtained for Texas. And just like for Texas, the main difference between the grade-level estimates in Panel B and the enrollment-cohort estimates in Panel A is that the enrollment-cohort approach yields a larger adverse effect of boys on boys and a smaller adverse effect of boys on girls. This is consistent with the selection bias of the grade-level approach in school systems with substantial retention (around 17% of the children in third grade in North Carolina have been retained in a lower grade).

Table 3 reports our findings for Italy. Panel A contains our results using the enrollment-cohort approach. Estimates are based on almost 1.6 million children in more than 8,000 schools. Recall that for Italy, we have to use test scores in second grade as third-grade test scores are unavailable. The effect of the share of boys in the enrollment cohort on girls in columns (1)-(4) is around -0.012 to -0.047. The effect on boys in columns (5)-(8) is -0.06 to -0.081. Comparing the same specifications with controls for predetermined characteristics yields adverse effects of boys on boys that are around 2-4 times the adverse effects of boys on girls. This difference is statistically significant at the 10-percent level.

Panel B of Table 3 reports results for Italy using the grade-level approach. The estimates are more similar to what we obtained using the enrollment-cohort approach when compared to our findings for Texas and North Carolina. For example, for Italy, both the enrollment-cohort approach and the grade-level approach yield a stronger adverse effect of boys on boys than on girls. This is consistent with the selection bias of the grade-level approach we highlight as there is much less grade retention in Italy than in Texas and North Carolina (less than 2% of the children in second grade in Italy have been retained in a lower grade). Also, the difference between the adverse effects of boys on boys and on girls with controls for predetermined characteristics for Italy is similar using the enrollment-cohort approach and the grade-level approach.

6 Conclusions

There is agreement in the literature that boys are more likely to cause trouble in school than girls and that this slows down learning when there are more boys in the classroom. This has been shown using within-school variation at a given grade level in Texas and Israel (Hoxby, 2000; Lavy and Schlosser, 2011) and within-school random assignment across classrooms in China (Hu, 2015; Gong et al., 2021).

Where findings diverge is in whether boys and girls are affected similarly or whether the trouble caused by boys has stronger adverse effects on boys. The approach using within-school grade-level variation yields similar adverse effects of boys on boys and girls, while within-school random assignment across classrooms yields that adverse effects on boys are 2-4 times stronger than adverse effect on girls. We argue that these divergent findings stem from a methodological issue. In school systems with grade retention, the approach based on within-school grade-level variation is subject to a simple selection bias. Under a condition that we find to be satisfied empirically, the selection bias implies that the true adverse effect of boys on the skills of boys is underestimated and the true adverse effect of boys on the skills of girls is overestimated. Hence, the selection bias may result in similar adverse effects of boys on how much boys and girls learn in school, although the adverse effect on boys is actually stronger than the adverse effect on girls.

We propose an alternative to the approach based on within-school grade-level variation. The alternative is an intention-to-treat approach based on variation in the share of boys and girls across enrollment cohorts as defined by the school system's primary-school enrollment policy. An important advantage of the enrollment-cohort approach compared to the grade-level approach, is that it is not subject to a selection bias in school systems with grade retention and will therefore not indicate peer effects of boys on boys or on girls if there are no peer effects. Moreover, in contrast to the grade-level approach, the enrollment-cohort approach will indicate positive peer effects if and only if peer effects are positive.

Our results using the enrollment-cohort approach confirm the adverse effect of boys on how much children learn in school. This is the case in Texas, North Carolina, and Italy. We find an adverse effect of boys on boys that is at least double the adverse effect on girls. This is consistent with the findings relying on within-school random assignment across classrooms. When we instead use the grade-level approach, results for the three school systems depend on the extent to which there is grade retention. In Texas and North Carolina, two school systems with substantial grade retention, the grade-level approach yields that boys have a similar adverse effect on how boys and girls learn in school. In Italy, where there is much less grade retention, the grade-level approach yields results that are similar to the enrollment-cohort approach. These differences across school systems are consistent with the selection bias of the grade-level approach we highlight.

References

Anelli, M. and G. Peri, "Peers' Composition Effects in the Choice of College Major," *The Economic Journal*, 2019, *129* (February), 553–602.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman, "School Quality and the Gender Gap in Educational Achievement," *American Economic Review*, 2016, *106* (5), 289–295.

____, ____, ____, ____, and ____, "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes," *American Economic Journal: Applied Economics*, 2019, *11* (3), 338–381.

Bach, M. and S. Sievert, "Birth Cohort Size Variation and the Estimation of Class Size Effects," *Journal of Human Resources*, 2024. Forthcoming.

Bedard, K. and E. Dhuey, "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *Quarterly Journal of Economics*, 2006, *121* (4), 1437–1472.

Bertrand, M. and J. Pan, "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal: Applied Economics*, 2013, *5* (1), 32–64.

Black, S. E., P. J. Devereux, and K. G. Salvanes, "Under Pressure? The Effect of Peers on Outcomes of Young Adults," *Journal of Labor Economics*, 2013, *31* (1), 119–153.

Bostwick, V. and B. A. Weinberg, "Nevertheless, She Persisted? Gender Peer Effects in Doctoral STEM Programs," *Journal of Labor Economics*, 2022, *40* (2), 397–436.

Brenøe, A. A. and U. Zølitz, "Exposure to More Female Peers Widens the Gender Gap in STEM Participation," *Journal of Labor Economics*, 2022, *38* (4), 1009–1054.

Cools, S. and E. Patacchini, "Peer Effects in Education," 2021, pp. 255–274.

Diette, T. M. and R. U. Oyelere, "Gender and Race Heterogeneity: The Impact of Students with Limited English on Native Students' Performance," *American Economic Review*, 2014, *104* (5), 412–417.

Epple, D. and R. E. Romano, "Peer Effects in Education: A Survey of the Theory and Evidence," 2011, *1A*, 1053–1163.

European Commission, "Grade Retention During Compulsory Education in Europe: Regulations and Statistics," 2021.

Gong, J., Y. Lu, and H. Song, "Gender Peer Effects on Students' Academic and Noncognitive Outcomes: Evidence and Mechanisms," *Journal of Human Resources*, 2021, *56* (3), 686–710.

Hill, A. J., "The Girl Next Door: The Effect of Opposite Gender Friends on High School Achievement," *American Economic Journal: Applied Economics*, 2015, *7* (3), 147–177.

Hoxby, C., "Peer Effects in the Classroom: Learning from Gender and Race Variation," Technical Report Working Paper 7867, National Bureau of Economic Research 2000.

——, "The Power of Peers," *Education Next*, 2002, *2* (2).

Hu, F., "Do Girl Peers Improve Your Academic Performance?," *Economics Letters*, 2015, *137*, 54–58.

Lavy, V. and A. Schlosser, "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, 2011, *3* (2), 1–33.

Lu, F. and M. L. Anderson, "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups," *Journal of Labor Economics*, 2015, *33* (1), 91–122.

Sacerdote, B., "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," 2011, *3*, 249–277.

——, "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?," *Annual Review of Economics*, 2014, *6*, 253–272.

Schneeweis, N. and M. Zweimüller, "Girls, Girls, Girls: Gender Composition and Female School Choice," *Economics of Education Review*, 2012, *31* (4), 482–500.

Schøne, P., I. Hardoy, and O. Raaum, "Peer Gender and Educational Choices," *Empirical Economics*, 2019, *59*, 1763–1797.

Texas Education Agency, "Grade Level Retention in Texas Public Schools," 1999. Office of Policy Planning and Research, also 2011.

Tyre, Peg, "The Trouble with Boys: A Surprising Report Card on Our Sons, Their Problems at School, and What Parents and Educators Must Do," 2008.

UNESCO, "Education for All: Is the World on Track?," 2002.

Whitmore, D., "Resource and Peer Impacts on Girls," *American Economic Review, Papers and Proceedings*, 2005, *95* (2), 199–203.

Main Tables

Table 1: Results for Third Grade in Texas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Enrollment-Cohort Approach								
	Girls				Boys			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Share of boys in enrollment cohort	-0.061**	-0.060*	-0.049*	-0.052	-0.127***	-0.122***	-0.125***	-0.121***
	(0.030)	(0.034)	(0.028)	(0.032)	(0.030)	(0.033)	(0.028)	(0.031)
Individual controls			yes	yes			yes	yes
Individuals	577507	577507	577507	577507	584663	584663	584663	584663
Schools	4319	4319	4319	4319	4349	4349	4349	4349
Panel B: Grade-Level Approach								
	Girls				Boys			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Share of boys in grade	-0.098***	-0.088**	-0.090***	-0.079**	-0.055*	-0.071*	-0.062*	-0.083**
	(0.033)	(0.039)	(0.031)	(0.037)	(0.033)	(0.039)	(0.030)	(0.035)
Individual controls			yes	yes			yes	yes
Individuals	584002	584002	584002	584002	588410	588410	588410	588410
Schools	4207	4207	4207	4207	4218	4218	4218	4218

Notes: The table contains estimates of the effect of the share of boys on the test scores of boys and girls in third grade in primary schools in Texas. Results are reported separately for mathematics and reading. Columns (1)-(4) contain the estimates of the effect of boys on girls' reading and mathematics test scores. Columns (5)-(8) contain the estimates of the effect of boys on boys' reading and mathematics test scores. Individual test scores are standardized using the average and standard deviation among students taking the same test. Panel A examines how the test scores of boys and girls in an enrollment cohort depend on the share of boys in the enrollment cohort using estimating equation (14). Enrollment cohorts are defined by the primary-school enrollment rule of the school system. Panel B examines how the test scores of boys and girls in third grade in the same school year depend on the share of boys in the grade using estimating equation (15). The grade-level results are for third grade in the school years starting 2005-2008. The enrollment-cohort results are for students who would have been in third grade in the school years starting 2005-2008 had they enrolled in primary school according to the rule in Texas and not been retained or skipped a grade in primary school. See the main text for details. Individual controls are race, ethnicity, and eligibility for a subsidized lunch. All specifications include school fixed effects. Estimates in Panel A include enrollment-cohort fixed effects and estimates in Panel B include school-year fixed effects. Standard errors reported in parentheses are clustered at the school level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level respectively.

Table 2: Results for Third Grade in North Carolina

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Enrollment-Cohort Approach								
	Girls				Boys			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Share of boys in enrollment cohort	-0.031	-0.062*	-0.033	-0.061*	-0.125***	-0.155***	-0.124***	-0.152***
	(0.031)	(0.034)	(0.029)	(0.033)	(0.032)	(0.035)	(0.030)	(0.033)
Individual controls			yes	yes			yes	yes
Individuals	428055	428055	428055	428055	438116	438116	438116	438116
Schools	1480	1480	1480	1480	1491	1491	1491	1491
Panel B: Grade-Level Approach								
	Girls				Boys			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Share of boys in grade	-0.053	-0.095**	-0.044	-0.082**	-0.026	-0.089**	-0.038	-0.101***
	(0.033)	(0.037)	(0.031)	(0.036)	(0.036)	(0.038)	(0.033)	(0.036)
Individual controls			yes	yes			yes	yes
Individuals	474996	474996	474996	474996	488944	488944	488944	488944
Schools	1447	1447	1447	1447	1454	1454	1454	1454

Notes: The table contains estimates of the effect of the share of boys on the test scores of boys and girls in third grade in primary schools in North Carolina. Results are reported separately for mathematics and reading. Columns (1)-(4) contain the estimates of the effect of boys on girls' reading and mathematics test scores. Columns (5)-(8) contain the estimates of the effect of boys on boys' reading and mathematics test scores. Individual test scores are standardized using the average and standard deviation among students taking the same test. Panel A examines how the test scores of boys and girls in an enrollment cohort depend on the share of boys in the enrollment cohort using estimating equation (14). Enrollment cohorts are defined by the primary-school enrollment rule of the school system. Panel B examines how the test scores of boys and girls in third grade in the same school year depend on the share of boys in the grade using estimating equation (15). The grade-level results are for third grade in the school years starting 1999-2008. The enrollment-cohort results are for students who would have been in third grade in the school years starting 1999-2008 had they enrolled in primary school according to the rule in North Carolina and not been retained or skipped a grade in primary school. See the main text for details. Individual controls are race and ethnicity. All specifications include school fixed effects. Estimates in Panel A include enrollment-cohort fixed effects and estimates in Panel B include school-year fixed effects. Standard errors reported in parentheses are clustered at the school level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level respectively.

Table 3: Results for Second Grade in Italy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Enrollment-Cohort Approach								
	Girls				Boys			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Share of boys in enrollment cohort	-0.016	-0.047	-0.012	-0.042	-0.060**	-0.083***	-0.061**	-0.081***
	(0.029)	(0.032)	(0.029)	(0.032)	(0.028)	(0.030)	(0.028)	(0.030)
Individual controls			yes	yes			yes	yes
Individuals	774543	774543	774543	774543	799277	799277	799277	799277
Schools	8032	8032	8032	8032	8024	8024	8024	8024
Panel B: Grade-Level Approach								
	Girls				Boys			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math
Share of boys in grade	-0.036	-0.060*	-0.032	-0.054	-0.069**	-0.091***	-0.072**	-0.091***
	(0.030)	(0.033)	(0.030)	(0.033)	(0.029)	(0.031)	(0.029)	(0.031)
Individual controls			yes	yes			yes	yes
Individuals	773283	773283	773283	773283	798332	798332	798332	798332
Schools	7881	7881	7881	7881	7884	7884	7884	7884

Notes: The table contains estimates of the effect of the share of boys on the test scores of boys and girls in second grade in primary schools in Italy. Results are reported separately for mathematics and reading. Columns (1)-(4) contain the estimates of the effect of boys on girls' reading and mathematics test scores. Columns (5)-(8) contain the estimates of the effect of boys on boys' reading and mathematics test scores. Individual test scores are standardized using the average and standard deviation among students taking the same test. Panel A examines how the test scores of boys and girls in an enrollment cohort depend on the share of boys in the enrollment cohort using estimating equation (14). Enrollment cohorts are defined by the primary-school enrollment rule of the school system. Panel B examines how the test scores of boys and girls in third grade in the same school year depend on the share of boys in the grade using estimating equation (15). The grade-level results are for second grade in the school years starting 2012-2016. The enrollment-cohort results are for students who would have been in second grade in the school years starting 2012-2016 had they enrolled in primary school according to the rule in Italy and not been retained or skipped a grade in primary school. See the main text for details. The individual controls are the educational attainment of mother and father (primary education, lower secondary education, upper secondary education, and college). All specifications include school fixed effects. Estimates in Panel A include enrollment-cohort fixed effects and estimates in Panel B include school-year fixed effects. Standard errors reported in parentheses are clustered at the school level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level respectively.

Appendix Table

Balancedness of Predetermined Skills at the Enrollment-Cohort Level

	(1)	(2)	(3)	(4)
	Results for Texas			
	Girls		Boys	
	Reading	Math	Reading	Math
Share of boys in enrollment cohort	-0.014	-0.010	-0.001	-0.000
	(0.009)	(0.008)	(0.008)	(0.008)
Individuals	577507	577507	584663	584663
	Results for North Carolina			
	Girls		Boys	
	Reading	Math	Reading	Math
Share of boys in enrollment cohort	0.002	-0.001	-0.002	-0.003
	(0.009)	(0.009)	(0.009)	(0.010)
Individuals	428055	428055	438116	438116
	Results for Italy			
	Girls		Boys	
	Reading	Math	Reading	Math
Share of boys in enrollment cohort	-0.006	-0.006	0.005	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Individuals	774543	774543	799277	799277

Notes: The table examines whether children's predetermined skills are balanced at the enrollment-cohort level in two steps. We first obtain our measure of children's predetermined skills by predicting their third-grade test scores based on their predetermined characteristics, school fixed effects, and enrollment-cohort fixed effects, separately for boys and girls and for reading and mathematics. To see whether the predetermined skills of children in an enrollment cohort depend on the share of boys in the enrollment cohort, we then regress the predicted third-grade test scores of children on the share of boys in their enrollment cohort. Columns (1)-(2) report estimates for the effect of the share of boys in the enrollment cohort on girls' predicted test scores and columns (3)-(4) report estimates for the effect of the share of boys in the enrollment cohort on boys' predicted test scores. All specifications include school fixed effects and enrollment-cohort fixed effects. Standard errors reported in parentheses are clustered at the school level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level respectively. Although estimates are quite precise, all estimates in the table are statistically insignificant, indicating that children's predetermined skills are balanced at the enrollment-cohort level.