

DISCUSSION PAPER SERIES

IZA DP No. 18208

**Freed from the Boys: How Single-Sex
Schooling Shapes Girls' Effort and
Performance in High-Stakes Exams**

Caterina Calsamiglia
Yarine Fawaz
Daniel Fernández-Kranz
Junhee Lee

OCTOBER 2025

DISCUSSION PAPER SERIES

IZA DP No. 18208

Freed from the Boys: How Single-Sex Schooling Shapes Girls' Effort and Performance in High-Stakes Exams

Caterina Calsamiglia

IPEG and ICREA

Yarine Fawaz

CEMFI and UAM

Daniel Fernández-Kranz

IE University and IZA

Junhee Lee

Korean Ministry of Foreign Affairs

OCTOBER 2025

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Freed from the Boys: How Single-Sex Schooling Shapes Girls' Effort and Performance in High-Stakes Exams

Prior research has found that boys often outperform girls in high-stakes math exams, raising the question of whether these gender differences under pressure stem from nature or nurture. This relative female disadvantage can influence access to selective university programs and subsequent career paths. Using administrative and survey data linked to a lottery-based school assignment system, we show that this disadvantage is reversed in single-sex schools: girls randomly assigned to SS schools devote more effort, outperform boys in high-stakes math exams, and have a higher likelihood of enrolling in university STEM degrees (excluding biology). These positive effects come at a cost to well-being in terms of higher stress and worse mental health. These effects are not driven by differences in teacher gender or school resources due to public versus private management. Our findings are consistent with theories emphasizing the social costs of norm violation: in single-sex schools, girls are freed from peer norms that may otherwise discourage overt academic ambition, allowing them to sustain higher effort in competitive and male-dominated domains.

JEL Classification: I21, J16, I24, D91, J24, I28

Keywords: gender, random assignment, single-sex schooling, nurture, education, high-stakes exams, Korea, gender gap, natural experiment

Corresponding author:

Daniel Fernández-Kranz

IE University

María de Molina, 31

28006 Madrid

Spain

E-mail: Daniel.Fernandez@ie.edu

1 Introduction

Gender differences in performance under pressure are a recurring finding in education research. In high-stakes math exams, boys often outperform girls, even when the same girls perform as well or better in low-stakes settings, i.e., a “high-stakes performance premium” for males (Azmat, Calsamiglia, and Iriberrí (2016), Schlosser, Neeman, and Attali (2019), Cai, Lu, Pan, and Zhong (2019)).¹

This pattern has raised a long-standing question: are such gender differences the product of nature or nurture? Do they reflect innate differences in reactions to stress, or are they shaped by the social and educational environments in which students develop? Understanding this distinction is crucial because it determines whether these differences are malleable. If nurture matters, then features of the learning environment—such as peer composition, classroom norms, or social expectations around competition—could influence how students perform when stakes are high. For instance, girls may face social costs when displaying overt ambition or competitiveness, which could dampen effort in mixed-gender settings. Conversely, environments that relax these social constraints might allow girls to perform to their full potential.

In this paper, we investigate the gender gap in performance under a particularly relevant high-stakes context—the CSAT university entrance exam—focusing on how it is shaped by the school peer environment, in particular, by single-sex versus coeducational schooling. Consistent with previous research (Dustmann, Ku, et al. (2018)), we find that single-sex schooling improves the educational outcomes of both boys and girls. Despite these gains, boys and girls respond very differently to the pressure associated with the final stage of their education before university. Girls in single-sex schools tend to self-assess more harshly, experience higher levels of stress, and devote greater effort to studying and attending tutoring sessions. Boys, in contrast, are barely affected or react in the opposite way. As a result of these divergent behavioral responses, girls in single-sex schools outperform their male counterparts (in relative terms) on the math component of the high-stakes CSAT exam, although this comes at the cost of increased stress and potential mental health issues. Single-sex schooling also increases the likelihood that girls enroll in STEM university degrees—excluding biology. Importantly, we show that these patterns are not explained by differences in school characteristics, beyond the absence/presence of peers of the opposite sex.

To the best of our knowledge, we are the first to show that the well-documented male performance premium in high-stakes math exams that determine university access reverses in single-sex schools. To show this reversal and its mechanisms, we draw on a unique setting that combines random assignment of students to school-type with detailed school and student-level administrative and survey data. We exploit the quasi-random assignment of students to single-sex versus coeducational schools under Korea’s Equalization Policy, which serves as our core source of identification. We combine this with administrative data that record students’ exam performance alongside detailed in-

¹interestingly for our study Schlosser et al. (2019) finds no performance gap between low and high stakes for Asians (its “no-gap” result referred to the quantitative section of the GRE).

formation on class composition, teacher gender, and school ownership, allowing us to rule out these observable school characteristics as alternative explanations. We further complement this with longitudinal survey evidence on students’ effort, tutoring, stress, and mental health, which enables us to uncover the behavioral mechanisms underlying the observed patterns.

The random assignment we exploit comes from Korea’s Equalization Policy, which allocated students to single-sex or coeducational high schools through district-level lotteries. This allocation mechanism was introduced under the Equalization Policy, which has been in effect since 1974 across all major metropolitan areas and in some rural regions.² Within this institutional framework, we examine how gender differences in academic performance vary between two national standardized exams: the College Scholastic Ability Test (CSAT, or “Suneung” in Korean), a high-stakes university entrance exam taken in the third and final year of high school, and the National Assessment of Educational Achievement (NAEA), a low-stakes exam administered in the second year of high school.

The CSAT represents the most consequential assessment in a Korean student’s academic career, as only those scoring in the top three percent nationally (with slight variation across years) gain admission to the elite SKY universities—Seoul National University, Korea University, and Yonsei University—which serve as gateways to the most prestigious career paths. In contrast, the NAEA is designed to monitor and evaluate educational quality at the national level and carries no implications for university admissions. Both exams are centrally administered by the Korea Institute for Curriculum and Evaluation (KICE), which oversees all aspects of exam development, administration, and scoring.

Our results are statistically and economically significant. We find that being randomly assigned to a single-sex school leads girls to close the math gender gap by nearly 3 standardized points in the high-stakes university entrance exam, a pivotal shift in a system where only the top 2-3% of students gain admission to elite SKY universities. We interpret our findings as broadly consistent with prior work showing that girls behave more like boys in competitive environments when educated in single-sex settings (A. Booth and Nolen (2012)). Also, and in line with theories emphasizing the social cost of norm violation (Fryer Jr and Torelli (2010); Bursztyn, Egorov, and Jensen (2019)), our results suggest that girls in coeducational environments may internalize behaviors that discourage overt displays of academic ambition—particularly in male-dominated and competitive domains.

We contribute to several strands of literature. The results of our study are consistent with a growing body of experimental research highlighting the role of *nurture* in shaping gender differences in the willingness to compete. For example, Gneezy, Leonard, and List (2009) and Andersen, Ertac, Gneezy, List, and Maximiano (2013) show that the

²This random assignment feature provides an ideally clean setting to identify the effects of single-sex schooling, as in for instance Dustmann et al. (2018); Hahn and Wang (2019a); Park, Behrman, and Choi (2013). Nevertheless, as put forward in J. Lee and Park (2017), the random assignment was not always purely random, e.g. a choice lottery program was introduced in Seoul in 2010, which is why we will only consider a sample period corresponding to the assignment (to the first year of high school) taking place before 2010.

commonly reported finding that men are more inclined to compete than women (e.g., Gneezy, Niederle, and Rustichini (2003), Gneezy and Rustichini (2004), Niederle and Vesterlund (2007), Iriberry and Rey-Biel (2019); see also the review in Niederle and Vesterlund (2011)) is reversed in a matrilineal society. They show that Maasai women in Tanzania are less competitive than men, while in the matrilineal Khasi society of India, women are more competitive than men. Related to this, evidence from recent studies shows that curricular content can shift gender norms and behavior in lasting ways. Hara and Rodriguez-Planas (2025), for Japan, and A. Booth, Fan, Meng, and Zhang (2019), for China, show that when students are exposed to less gender-stereotypical content in school, girls adopt less traditional attitudes and behaviors later in life. Also on gender norms, Garcia-Brazales (2025) finds that being in primary school classes with a higher share of females weakens traditional gender views both for males and females, which translates into actual behavior decades later. Although we do not directly measure gender norms or willingness to compete, our results suggest that girls in single-sex schools display behavioral patterns plausibly associated with greater ambition and a stronger ability to perform under pressure. These results are consistent with Buser, van den Assem, and van Dolder (2023) and A. Booth and Nolen (2012), who find that girls are more willing to compete - and behave more like boys in this regard - when competing against other girls rather than against boys.

We also contribute to the literature on how peer gender composition affects school performance, where evidence is mixed. Some studies show that a higher proportion of girls enhances female academic performance (Lavy and Schlosser (2011); Eisenkopf, Hესsami, Fischbacher, and Ursprung (2015); A. L. Booth, Cardona-Sosa, and Nolen (2018)), while others find no effect (S. Lee, Turner, Woo, and Kim (2014); Doris, O’Neill, and Sweetman (2013)). However, none of these studies examine whether peer composition influences the gender gap in high-stakes exams.

Our findings on single-sex schooling and gender differences in mathematics performance contribute to the ongoing debate initiated by Fryer Jr and Levitt (2010). In their critique of Guiso, Monte, Sapienza, and Zingales (2008), Fryer and Levitt show that the negative cross-country correlation between gender gaps in mathematics and measures of gender equality is not robust to the inclusion of a group of Muslim-majority countries—such as Iran, Jordan, and Bahrain—where girls outperform boys in math despite these countries ranking low on standard gender equality indices. They argue that what these countries have in common is the prevalence of sex-segregated schooling environments. Once such countries (including Korea) are excluded from the sample, a positive relationship between gender equality and female relative performance in mathematics reemerges. Our study adds nuance to this discussion by showing that, within Korea, the association between single-sex schooling and the gender gap in mathematics is more complex than suggested by cross-country correlations. Specifically, we find that gender gaps in math performance—measured in both low- and high-stakes assessments—are more pronounced among students attending single-sex schools than among those in co-educational schools. However, this gap narrows substantially, and in some cases even reverses, in high-stakes settings. These findings suggest that while single-sex school-

ing may be associated with larger average gender gaps, it may also foster resilience or improved performance under pressure among female students.

Last but not least, our study also connects to the broader literature on fertility decline in high-income countries, with South Korea representing one of the most extreme cases. Recent work by [Kearney and Levine \(2025\)](#) highlights how persistent economic pressures and evolving social norms jointly explain why fertility is declining and has remained so low despite supportive policy interventions. [Kleven \(2022\)](#) shows that gender norms strongly mediate the large “child penalties” women face after childbirth, shaping both fertility decisions and long-run gender inequality. By studying how the school peer environment influences the development of gendered responses to pressure, we contribute to understanding an upstream mechanism through which gender norms are formed—mechanisms that are ultimately central for family formation and fertility outcomes.

2 Institutional Background and Data

2.1 The High School Equalization Policy

The lottery-based assignment of students to high schools within districts in South Korea has been well documented in prior research, e.g. [Choi, Moon, and Ridder \(2014\)](#); [Dustmann et al. \(2018\)](#); [Park et al. \(2013\)](#); [Wang \(2015\)](#). The origins of the High School Equalization Policy (HSEP) lottery date back to 1974.³ All the metropolitan areas of Korea, starting with Seoul and Busan, gradually adopted the HSEP, whose ultimate purpose was to reduce educational inequality and competition among middle school students applying to high school. Before HSEP, South Korea was grappling with an education system marked by intense competition and stark inequalities. Middle school students, barely in their teens, were subjected to grueling entrance examinations in order to secure a spot in elite academic high schools. The pressure to succeed led to a proliferation of private tutoring academies (*hagwon*) and a widening gap between students who could afford extra preparation and those who could not. Educational outcomes—and, by extension, future life chances—were increasingly determined by a family’s socioeconomic status. In response, the South Korean government launched the Equalization Policy in 1974, beginning with the city of Seoul. The aim was to level the playing field and reduce the extreme academic competition that had come to define the high school admission process. The policy eliminated competitive entrance exams for most general high schools in designated cities and instead introduced a lottery-based school assignment system within each district. This randomization was designed to ensure a more equitable distribution of students and to discourage the sorting of children into ‘elite’ and ‘non-elite’ institutions. The government hoped that by reducing the stakes of high school placement, it could also reduce the reliance on private tutoring and mitigate the psychological and financial burdens placed on families. Over the years, the policy was

³The HSEP was first enforced at the middle-school level, starting in 1968. Enforcement was even more systematic than for high schools.

gradually expanded to other metropolitan areas Gwangju. By the mid-2000s, all middle schools and around 70% of high schools were under the HSEP regime. The policy was not enforced in rural areas and in the areas where it was enforced, there was still a small number of special-purpose schools, specialized in athletics or arts for instance, or vocational schools, which could escape the random assignment of students. Roughly 5 percent of all high school students are enrolled in these schools.

In areas where the HSEP was rigorously implemented—such as Seoul and many other metropolitan regions—school districts tend to be quite large. Seoul, for example, is divided into just 11 districts despite having a population of over 10 million. As a result, each district includes, on average, 14 high schools only for boys and 13 only for girls, in addition to coeducational schools.

This institutional setup offers a unique opportunity to explore whether gender differences in response to pressure vary systematically with school type. Since students are allocated to high schools within their district independently of whether the school is single-sex or coeducational, we can examine this question without the usual concerns about endogenous sorting. Rather than revisiting the well-studied effects of single-sex schooling on academic outcomes (see, e.g., [Park et al. \(2013\)](#); [Dustmann et al. \(2018\)](#)), we use this context to ask whether the presence or absence of gender gaps under high-pressure conditions is shaped by the school environment itself.

This analysis is further strengthened by the broader policy goals of the Equalization Policy, which aimed to standardize key aspects of schooling across the country ([Hahn, Wang, and Yang \(2018\)](#)). All high schools—regardless of type—follow a centrally designed national curriculum, operate on the same academic calendar, and are subject to uniform teacher qualification standards and compensation schemes. Moreover, school funding is equalized through centralized financing, with even private schools receiving government subsidies. This high degree of institutional uniformity helps ensure that any observed differences in student outcomes are less likely to reflect variation in school quality, allowing us to focus more precisely on differences in peer composition and school environment.

While the Equalization Policy ensures a high degree of uniformity across schools, this does not imply that single-sex and coeducational schools are identical in all respects. Notably, single-sex schools are far more likely to be private: according to our data, approximately 70% of single-sex schools are private, compared to just 24% of coeducational schools (see Table [A.1](#)). However, under the policy, private and public schools operate under nearly identical conditions. In return for substantial government subsidies, private schools are required to adopt the national curriculum, charge the same tuition as public schools, and participate in the centralized student assignment process. Teachers in private schools are also paid according to the government salary schedule. The primary distinction is that private schools retain autonomy over personnel decisions, such as hiring. As a result, private schools in Korea function as de facto quasi-public institutions. Nonetheless, in our analysis we control for school ownership (public vs. private) and find that our results are unaffected by this distinction.

Until 2010, when the Equalization Policy in Seoul was replaced by a more flexible

school-choice lottery system.⁴ compliance with the random assignment mechanism was virtually complete. The only way for a student to avoid being assigned to an undesired school was for their family to change residence to a different school district—and this move had to occur at least one year prior to assignment. Even then, families relocating to a new district would still be subject to random assignment within that district. In other words, the only form of choice available to families was residential sorting.

As documented by Kwak, Ku, et al. (2013), there is no evidence—anecdotal or otherwise—of corruption or manipulation in the high-school lottery system. No legal cases, media reports, or social media accounts suggest that families were able to circumvent the randomization process. As Kwak et al. (2013) notes, “The inability to influence the school assignment—beyond residential sorting—is quite consistent with the strong sentiment shared among South Korean parents toward the education of their children; had any parents learned about another parent’s ‘successful’ manipulation of the high school lottery, they would have immediately protested and brought the case to the attention of the media and to the court.”

Although our administrative data covers the years 2010 and 2011—during the period when the HSEP was being phased out—the students we observe taking the NAEA and CSAT exams were still strictly subject to the original random assignment rules. The reform only affected students entering high school from 2010 onward, which means the first cohort not fully subject to the EP began their senior year (and took the CSAT) in 2012.

2.2 The Korean high school education system and key national assessments

South Korea’s high school education (Grades 10–12) sits atop the 6-3-3 ladder and is formally divided into two broad school types. Roughly three-quarters of students attend “academic” general high schools that feed the university track, while about one-quarter enroll in vocational/technical schools oriented toward the labour market. Inside the academic high schools, a second layer of differentiation emerges beginning in Grade 11: pupils sort into a liberal-arts/social-sciences stream (mun-gwa) or a natural-sciences stream (i-gwa), the latter featuring heavier doses of advanced mathematics, physics, chemistry and biology, whereas the former emphasises history, geography, economics and a lighter mathematics syllabus. Track allocation is typically managed within the same school building; students share the common Grade-10 curriculum and then declare their preferred stream through school counselling procedures at the end of the first high-school year.

Although the ministry’s 7th National Curriculum (1997, phased into high schools in 2003) enlarged the menu of electives, it preserved the basic humanities–science di-

⁴Under the new system, students could apply to any high school across the city rather than being restricted to their residential district. Admissions occurred in multiple rounds: a city-wide lottery for students’ top two school choices (accounting for roughly 20–60% of available seats), followed by a district-level lottery for the next two preferences (about 40% of seats), and a final round assigning remaining students based on factors such as commute time or religion.

chotomy. Once students select a track in Grade 11 all students in the same track take the same curriculum, with only limited freedom to mix courses across streams. This structure is reinforced by the College Scholastic Ability Test (CSAT), which requires science-track examinees to sit Calculus and at least two science subjects, while humanities-track students combine an easier mathematics paper with social-studies options. Because CSAT subject bundles are hard-wired into university admissions criteria, switching tracks after Grade 11 is legally possible but academically self-defeating and therefore exceedingly rare; students who attempt it would face significant curricular gaps and a mis-match with CSAT preparation schedules

In June of the second year of high school (Grade 11, typically age 17), students take the National Assessment of Educational Achievement (NAEA), a standardized diagnostic test administered by the Korea Institute for Curriculum and Evaluation (KICE). It assesses academic achievement in Korean, Mathematics, and English, with additional subjects such as Science or Social Studies included for sampled groups. The exam is used primarily for national benchmarking and school-level performance monitoring; it does not influence university admissions and is not used for individual student selection. However, because it takes place after students have chosen their academic track, the NAEA provides a national snapshot of student learning within each track.

By contrast, the College Scholastic Ability Test (CSAT), or Suneung, is taken in November of the third and final year of high school (Grade 12), roughly 17 months after the NAEA. The CSAT plays a central role in determining university admission and is the culmination of years of preparation aligned with each student’s chosen track. Students are required to take Korean, Mathematics, and English, alongside two elective subjects that correspond to their track—either from the natural or social sciences. Universities weigh these subjects differently depending on the field of study; for example, engineering departments tend to assign greater weight to mathematics and science, while humanities departments emphasize Korean and English. The structure and timing of these two national exams—NAEA in June of Grade 11 and CSAT in November of Grade 12—reinforce the significance of early tracking in the Korean education system, with limited opportunity for reversals or course correction once the academic trajectory has been set.

3 Data

We combine administrative and longitudinal survey data. First, we use administrative data to document the presence of a gender gap in performance associated with a shift in stakes and to estimate the causal effect of single-sex schooling on gender-specific responses to high-stakes exams. Second, we use survey data to examine potential mechanisms through which single-sex schooling may influence students’ behavior and help explain our results.

3.1 Administrative data

Our administrative data combines student-level and school-level information. The student-level data include CSAT and NAEA records from 2010 to 2016, provided by the Korean Ministry of Education, Science and Technology. These were merged with school-level data from the EduData Service System (EDSS), obtained upon review and approval of our research proposal. The EDSS has made data publicly available since 2010, using a stratified sampling strategy that selects approximately 70 percent of schools nationwide. The dataset provides a rich array of school-level information, including school expenditures and teacher characteristics. We match the 2010 NAEA data with 2011 CSAT scores, forming a clean cohort unaffected by subsequent policy reforms. We focus on this cohort because, beginning in 2010, the Equalization Policy was substantially relaxed in Seoul through the introduction of a more flexible high school choice system. As a result, using CSAT data from 2012 onward would require excluding Seoul (for the 2012 graduating cohort) and an increasing number of regions in subsequent years, undermining comparability and complicating interpretation. We therefore base our results on the last cohort for which nationwide random assignment remained in place.

Besides test scores, the CSAT and NAEA data include individual-level information on each test taker, such as gender, a school identifier (common to the CSAT and NAEA data), and city and district information. Based on the school ID, we match students' performance at NAEA and CSAT in all three subjects (Korean, Maths, and English, which are common to the two exams and all exam takers) with school characteristics, including the gender type of school (coed vs. single-sex). We fix school characteristics—e.g. year of creation of establishment, establishment type (public versus private, general versus specialized), school size (total enrollment in each grade), class size, percentage of female teachers, school equipment, length of school year, share of unionized teachers, - to the year each cohort takes the CSAT exam.

To ensure that high school assignment within a district is as-good-as random, we restrict the sample to general academic high schools located in areas covered by the Equalization Policy. The sample is made of 373,229 observations belonging to 617 schools in year 2011 (with corresponding CSAT takers in 2011 and NAEA takers in 2010), 45 % of which are coed, 26% all-boys schools, and 28 % all-girls schools (see Figure A.1 for more details about the distribution of the share of females across schools).

Table A.1 shows descriptive statistics at the school level for coed versus single-sex schools. Coed and SS schools are quite similar in terms of class size, pupils-to-teacher ratio, duration of school year, equipment, or building area, but they also seem to differ across other dimensions. First, all-girls schools have a much higher share of female teachers (53%) than all-boys schools (29%), but still less than coed schools (59%). SS schools are on average much older than coed schools, and also more likely to be private. In terms of test scores, SS schools outperform coed schools across all subjects both for boys and girls. Test scores are standardized to have a mean of 100 and a standard deviation of 20 in each year. CSAT scores are already reported in this format, while NAEA scores are rescaled accordingly to enable comparability across exams. We also present the results in terms of percentile ranks.

Figure A.2 shows the distribution of female and male performance for all three subjects at the high and low-stakes exams. While in the US or Spanish context (Azmat et al. (2016) and Schlosser et al. (2019)), girls do better across all subjects, it seems Korea still exhibits a gender gap in Maths (similarly to Cai et al. (2019) for Science), which we will study in a formal empirical framework.

Table 1: Students' Performance: Single-sex versus coed.

Variable	Single-sex		Coeducational		All	
	SS boys	SS girls	Coed (boys)	Coed (girls)	Boys	Girls
<i>PANEL A: NAEA (Low Stakes) score</i>						
Math score	105.02***	99.88***	97.37	96.61	101.49	98.50
Korean score	99.02***	105.66***	91.94	102.45	95.76	104.30
English score	101.06***	103.24***	94.15	100.71	97.87	102.17
Observations	45896	48383	39364	35500	85260	83883
<i>PANEL B: CSAT (High Stakes) score</i>						
Math score	102.81***	100.63***	98.09	97.11	100.73	99.17
Korean score	100.54***	102.97***	95.56	100.22	98.33	101.83
English score	101.29***	102.25***	96.26	99.27	99.07	101.02
Observations	59345	56898	47343	40500	106688	97398

Stars indicate statistical significance of Welch two-sample t -tests comparing single-sex and coeducational schools.

SS boys are compared to coed boys, and SS girls are compared to coed girls.

The unit of analysis in this table is a student.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A first look at the raw, unadjusted test scores in Table [1](#) provides a glimpse into the central finding of our paper. Focusing on mathematics, we can observe how the gender gap evolves differently across school types when moving from the low-stakes NAEA to the high-stakes CSAT. In coeducational schools, the gender gap in math is initially small at low stakes, with boys scoring 97.49 and girls scoring 96.98, a difference of just 0.51 points. In the high-stakes exam, this gap widens slightly, with boys scoring 98.09 and girls 97.11, for a larger difference of 0.98 points. This pattern, where the male advantage in math increases under pressure, aligns with findings in some of the existing literature.

In sharp contrast, single-sex schools exhibit a completely different dynamic. At low stakes, there is a very large gender gap in favor of boys: SS boys score 104.81, while SS girls score 99.81, a substantial gap of 5.0 points. However, when moving to the high-stakes CSAT, this gap narrows dramatically. SS boys' scores fall to 102.81, while SS girls' scores rise to 100.63, reducing the gap to just 2.18 points. We will later corroborate this pattern in our regression-based empirical analysis.^{[5](#)}

3.2 Survey data

Our survey data come from the Korean Youth Panel Survey (KYPS). The KYPS is a nationally representative longitudinal dataset collected by the National Youth Policy Institute (NYPI) to study the development and life trajectories of Korean adolescents. It follows students from early adolescence through young adulthood, gathering annual data on academic performance, school life, family background, time use, extracurricular activities, mental and physical health, and transitions into higher education and the labor market. The survey uses a stratified multistage cluster sampling design, selecting schools and then randomly sampling an entire class within each school. Multiple cohorts have been launched since 2003. The younger cohort is first observed in 4th grade of elementary school and followed through 2nd year of middle school. The older cohort, which we focus on, consists of students aged 14 in 2003 (2nd year of middle school), who are tracked through high school and into university by 2008. For our purposes, we use data from the point of random high school assignment in 2004 until students take the CSAT in their final year of high school in 2007. Our KYPS cohort of interest—the older cohort—comprises 11,671 observations corresponding to 2,538 students residing in areas covered by the Equalization Policy.

KYPS offers two features particularly valuable to our study. First, it records students' residential districts, allowing us to assign them to their corresponding school districts—within which high school assignment is random in areas covered by the Equalization Policy. For instance, Seoul comprises 25 residential districts mapped to 11 school districts. We restrict our analysis to students in regions where the policy applied, covering roughly 70% of our sample. Second, KYPS provides rich information on family background, academic investment (e.g., study time, private tutoring), and students' attitudes toward school, family, peers, and broader aspects of adolescent life, including

⁵Because scores are standardized by subject and stakes, the decline in boys' math scores in SS under high stakes does not necessarily reflect a drop in their **absolute** performance, but instead a decline in their **relative** standing—or, equivalently, an improvement in the relative performance of other groups.

their health. Although students are asked to report their national percentile ranks on the NAEA and CSAT, we interpret these responses cautiously due to substantial non-response.

The KYPS dataset does not directly report the gender type of schools, but we recover this information using school identifiers and the gender composition of enrolled students. To construct a conservative classification, we pool data across years and define a school as “female-only” if all observed students are female, and “male-only” if all are male. Schools with both male and female students are classified as coeducational. In the first wave of the panel, the sampling design ensures accurate school-level gender composition: when a school is selected, an entire class is surveyed, and those same students are followed in subsequent years, regardless of class reassignment. In later waves, we observe between 1 and 96 students per school. To reduce miss-classification risk, we exclude schools with fewer than six observed students, resulting in the removal of approximately 15% of the sample.

Balance Tests A crucial assumption for our identification strategy is that the lottery-based assignment of students to either single-sex or coeducational schools is as-good-as random. To verify this, we conduct balance tests using the pre-high school data from the KYPS⁶. We compare a range of student, family, and academic background variables measured in the year prior to high school entry.

We first present simple comparisons of means in Table A.4 for girls and Table A.6 for boys. Without accounting for the randomization strata, these tables show some small but statistically significant differences. For instance, Table A.4 indicates that girls assigned to single-sex schools have slightly lower household income (-0.11 log points, $p < 0.05$) and fathers with less education (-0.46 years, $p < 0.10$). Similarly, Table A.6 shows that boys assigned to single-sex schools also have lower household income and paternal education.

However, the appropriate test of our research design compares characteristics *within* the randomization strata, which are the district-year cells. Tables A.5 and A.7 present these more rigorous tests by including district fixed effects in the regressions.

Once we condition on the district, the results strongly support the validity of the random assignment. As shown in Table A.5 for girls, the previously significant differences in family background become small and statistically insignificant. In fact, **none** of the eleven covariates show a statistically significant difference at conventional levels between girls assigned to single-sex versus coeducational schools.

The results for boys, presented in Table A.7, are similarly compelling. After including district fixed effects, ten of the eleven covariates show no significant differences. The only exception is a marginal difference in the likelihood of the father working full-time (-0.05, $p < 0.10$).

Of particular importance is the covariate *average performance*, which is a student’s self-reported academic standing prior to high school, expressed on a 1-5 scale (where

⁶Similar pre-high school data is not available in the administrative data.

1 is very poor, 2 is poor, 3 is average, 4 is good and 5 is very good.)⁷ For both boys (Table A.7) and girls (Table A.5), the difference in this variable between those assigned to single-sex versus coeducational schools is very small and not statistically significant once district fixed effects are included. This suggests that, on average, students entering both types of schools were observationally similar in terms of prior academic performance.

Overall, these balance tests provide strong support for our key identifying assumption. The assignment to a single-sex or coeducational school is not systematically correlated with a wide range of observable, pre-determined student characteristics within the relevant lottery pools. This allows us to attribute the differences in outcomes we observe to the causal effect of the school environment itself.

4 Empirical Strategy

A large body of prior research has documented that students in single-sex schools tend to outperform their peers in coeducational schools (Dustmann et al. (2018); Hahn and Wang (2019b); J. Lee and Park (2017))⁸ While we confirm this well-known performance premium in our data, our contribution lies in investigating whether the response to pressure differs systematically by gender and school type. In particular, we examine whether single-sex schooling shapes how boys and girls adjust their effort and performance when moving from a low-stakes to a high-stakes environment.

We begin by establishing the baseline gender gap in academic performance in our sample by estimating the following equation:

$$\text{Score}_{i,e} = \alpha_0 + \alpha_1 \text{Female}_i + \delta_{d(i),t} + \epsilon_{i,e} \quad (1)$$

where $\text{Score}_{i,e}$ is the academic performance of student i on exam e (either the low-stakes NAEA or the high-stakes CSAT). The variable Female_i is an indicator equal to one if the student is female. The coefficient of interest, α_1 , captures the average performance difference between female and male students across both exams. Crucially, $\delta_{d(i),t}$ represents district-year fixed effects, which absorb any time-invariant differences across school districts as well as any year-specific shocks common to all students within a district. This ensures our estimates are based on comparisons of students who were subject to the same local lottery pool.

Next, we move to our first key question: do gender differences in performance change when the stakes of the exam increase? We adopt a difference-in-differences (DiD) framework to estimate the differential response of girls versus boys to high-stakes testing environments. The model is specified as:

⁷Non-response for this variable affects more than 40% of observations, hence this result must be taken cautiously.

⁸J. Lee and Park (2017) finds that, when given a choice, students exhibit a clear preference for assignment to single-sex schools. Dustmann et al. (2018) attempts to unpack this performance gap, and finds that whereas boys performance is affected by the gender of peers at the school-level, it is the gender of class-level peers that affects girls. Hahn and Wang (2019b) further points to differences in time use and participation in extracurricular activities as contributing factors

$$\text{Score}_{i,e} = \alpha_0 + \alpha_1 \text{Female}_i + \alpha_2 \text{High stakes}_e + \alpha_3 (\text{Female}_i \times \text{High stakes}_e) + \delta_{d(i),t} + \epsilon_{i,e} \quad (2)$$

Here, High stakes_e is a dummy variable equal to 1 for the high-stakes CSAT and 0 for the low-stakes NAEA. The DiD coefficient, α_3 , is our primary interest in this model. It measures how the gender gap changes when moving from the low-stakes to the high-stakes exam. A negative α_3 , as found in much of the Western literature (e.g., Azmat et al. (2016), Schlosser et al. (2019), and Cai et al. (2019)), would imply that girls' performance declines relative to boys' under pressure. Since our administrative dataset links students at the school level rather than providing a unique student identifier across both exams, this model compares performance across cohorts within the same schools and districts. To ensure robustness, the dependent variable, Score is defined in two ways: as a percentile rank and a standardized score (mean 100, standard deviation 20).

The central analysis of our paper extends this model to a difference-in-difference-in-differences (DDD) framework to test whether single-sex schooling is the key moderator of these effects. The quasi-random assignment of students to either single-sex or coeducational schools allows us to cleanly identify this third-level interaction. Our full model is:

$$\begin{aligned} \text{Score}_{i,e} = & \alpha_0 + \alpha_1 \text{Female}_i + \alpha_2 \text{High stakes}_e + \alpha_3 (\text{Female}_i \times \text{High stakes}_e) \\ & + \alpha_4 \text{Single sex}_{s(i)} + \alpha_5 (\text{Female}_i \times \text{Single sex}_{s(i)}) \\ & + \alpha_6 (\text{High stakes}_e \times \text{Single sex}_{s(i)}) \\ & + \alpha_7 (\text{Female}_i \times \text{Single sex}_{s(i)} \times \text{High stakes}_e) + \delta_{d(i),t} + \epsilon_{i,e} \end{aligned} \quad (3)$$

It is important to note how this specification relates to prior literature. The well-documented “single-sex school premium” (e.g., Dustmann et al. (2018); Park et al. (2013)) is captured directly by coefficients in our model. The parameter α_4 estimates the average performance advantage for boys in single-sex schools relative to boys in coeducational schools (the baseline group). The corresponding advantage for girls is given by the sum $\alpha_4 + \alpha_5$. We test and corroborate this general finding in our own data. However, our paper's contribution goes beyond confirming this average effect. We focus on the higher-order interaction terms—specifically α_6 and our main coefficient of interest, α_7 —to understand how single-sex schooling modulates performance *specifically in response to pressure*. Thus, our analysis estimates the effect of single-sex schooling on the high-stakes versus low-stakes performance gap, *on top of* the average positive effect this type of schooling has on academic outcomes.

The coefficient on the triple interaction, α_7 , is our main parameter of interest. It captures whether the differential effect of high stakes on girls (measured by α_3) is *itself* different for girls in single-sex schools compared to those in coeducational schools. A significant α_7 would indicate that the single-sex environment fundamentally alters how girls respond to high-stakes academic pressure relative to boys and relative to their peers in coed schools. While this is conceptually similar to splitting the sample by gender and

school type, the unified DDD framework allows for a direct and statistically rigorous comparison of all effects within a single specification.

5 Results

In this section, we present the results from our empirical analysis. We begin by establishing the baseline gender gaps in performance across different subjects. We then show how these gaps evolve under high-stakes pressure and, finally, demonstrate the central role of single-sex schooling in driving these dynamics. For the sake of brevity and given its importance for future educational and labor market outcomes, we focus our detailed discussion on mathematics, where a significant gender gap persists in many countries, including Korea.

5.1 Gender Differences in Performance across Subjects

We first estimate the average performance difference between boys and girls using Equation (1). Table 2 presents the estimates of the coefficient α_1 (*Female*) for our three subjects of interest. These estimates, which are more rigorous than simple descriptive statistics as they incorporate district-year fixed effects, reveal a clear and differential pattern across subjects.

Table 2: Gender gap in performance across subjects

	Percentile Rank			Standardized Score		
	(1) Math	(2) Korean	(3) English	(4) Math	(5) Korean	(6) English
Female	-3.298*** (0.372)	7.701*** (0.346)	3.491*** (0.392)	-2.280*** (0.261)	5.712*** (0.273)	2.923*** (0.282)
Constant	45.337*** (0.179)	40.116*** (0.168)	42.150*** (0.190)	101.102*** (0.126)	97.223*** (0.133)	98.580*** (0.137)
r ²	0.02	0.04	0.04	0.02	0.04	0.04
N	363428	373022	370934	363428	373022	370934

All regressions include district-year fixed effects.

Standard errors in parentheses, clustered at district-year level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results in columns (1) to (3), which use percentile ranks as the outcome, are particularly revealing given that admission to the elite SKY universities is a matter of rank. In mathematics (column 1), girls rank, on average, 3.3 percentile points lower than boys. This is a substantial gap in a system where only the top 2-3% of students gain admission to the most prestigious institutions. The results using standardized scores in column (4) confirm this, showing a disadvantage of 2.28 points ($\alpha_1 = -2.280, p < 0.01$).

Conversely, girls significantly outperform boys in language subjects. The female advantage is largest in Korean, where girls rank 7.7 percentile points higher than boys ($\alpha_1 = 7.701, p < 0.01$). In English, the advantage is 3.5 percentile points ($\alpha_1 = 3.491, p <$

0.01). The standardized score results in columns (5) and (6) mirror these findings. Given that the math gap is often linked to disparities in STEM fields and represents the only subject where girls lag behind, we focus our subsequent analysis on understanding the factors that influence this particular gap. We will display the remainder of the findings in terms of standardized scores, in order not to overload tables.

5.2 The Differential Effect of High Stakes on Gender Gaps

We now turn to the core puzzle motivating this paper: how does this gender gap in mathematics change when the stakes are raised? We estimate the difference-in-differences model from Equation (2), with the results for mathematics presented in Table 3. Our coefficient of interest here is α_3 , on the interaction term *Female X High-stakes exam*. The table shows how our estimates evolve as we progressively add controls, demonstrating the robustness of our findings.

Table 3: Gender gaps in response to increased stakes in Math

	(1)	(2)	(3)	(4)	(5)
Female	-2.280*** (0.261)	-2.284*** (0.265)	-3.059*** (0.358)	-2.153*** (0.275)	-2.206*** (0.221)
High-stakes exam		-0.129 (0.403)	-0.831 (0.496)	-0.870* (0.507)	-1.747*** (0.225)
Female X High-stakes exam			1.451*** (0.298)	1.490*** (0.302)	1.531*** (0.254)
FE	District	District	District	District	School
Controls	No	No	No	School char.	School char.
R2	0.02	0.02	0.02	0.03	0.14
N	363428	363428	363428	361689	361689

Dep. var. is score rank at either CSAT (HS) or NAEA (LS) exam.

Standard errors in parentheses, clustered at level of FE.

School controls include: date of school creation, private vs public, share female peers and share female teachers.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (1) replicates our baseline finding from Table 2, showing an average math disadvantage of 2.28 points for girls. Column (2) introduces the *High-stakes exam* dummy, which has a small and insignificant coefficient, suggesting no major average performance shift between the NAEA and CSAT, on average for both genders.

The key result emerges in column (3), which includes the crucial interaction term *Female X High-stakes exam*. Here, we find a large, positive, and statistically significant coefficient on the interaction ($\alpha_3 = 1.451, p < 0.01$). This indicates that the initial 3.06-point disadvantage for girls in the low-stakes exam ($\alpha_1 = -3.059$) narrows dramatically in the high-stakes exam, a finding that runs contrary to much of the existing literature.

In columns (4) and (5), we test the robustness of this result by controlling for potential confounders. In column (4), we add a rich set of school-level characteristics, including the school's founding date, private vs. public status, and the share of female peers and teachers. The inclusion of these controls leaves our coefficient of interest,

α_3 , virtually unchanged ($\alpha_3 = 1.490, p < 0.01$), suggesting that these observable school features do not drive the differential performance of girls under pressure.

Finally, in column (5), we replace the district-level fixed effects and school controls with much more stringent school-level fixed effects. This specification absorbs all time-invariant differences between schools, ensuring that our estimate is based on comparing changes over time within the very same schools. Even under this highly demanding specification, the interaction coefficient remains positive and highly significant ($\alpha_3 = 1.531, p < 0.01$). The stability of this coefficient across all specifications provides strong evidence that female students in Korea genuinely perform better relative to boys when faced with a high-stakes exam. This surprising resilience motivates our investigation into its primary driver: the single-sex school environment.

Note that our findings echo some of the secondary results of previous literature. For instance, when looking at gender differences among the Asian subsample of GRE takers in the US, Schlosser et al. (2019) finds that the main result (a wider gap between low and high stakes for boys) does not extend to the Asian subsample (possibly due to insufficient power). In addition, the gender difference in performance between the Gaokao and a low-stakes exam found in Cai et al. (2019) is driven by non-mathematical subjects, and is barely significant when focusing on Maths. Our finding that girls do not “choke under pressure” in Maths, in the Korean context, suggests that the traditional “girls choking under pressure” phenomenon is not set in stone, and deserves further exploration. Particularly, given the dominance of single-sex schooling in Korea (57 % of our sample study in a single-sex school), we now turn to investigate the role of SS schooling in that finding.

5.3 The Role of Single-Sex Schooling: A Triple-Difference Analysis

Having established that girls exhibit remarkable resilience to pressure, we now investigate whether this phenomenon is driven by the school environment, specifically single-sex schooling. To do so, we estimate the full difference-in-difference-in-differences (DDD) model from Equation (3). The results, presented in Table 4, demonstrate step-by-step how single-sex schooling contributes to explaining the observed pattern.

The first three columns of Table 4 build the model by adding one key variable at a time. Columns (1) and (2) reproduce the results of Table 3. Column (3) introduces the main effect of single-sex schooling (*Single-sex school*). The coefficient α_4 is positive and highly significant ($\alpha_4 = 5.367, p < 0.01$), confirming the well-documented “single-sex premium”: on average, students (of both genders) in single-sex schools outperform students in coed schools by over 5 standardized points.

Column (5) introduces all the two-way interactions. The coefficient on *Female X High-stakes exam* (α_3) remains positive and significant ($\alpha_3 = 1.595, p < 0.01$), similar to our findings in Table 3. This shows that, before accounting for the full set of interactions, the overall trend of girls performing better under pressure still holds. The coefficient on *Female X Single-sex* (α_5) is large and negative ($\alpha_5 = -2.641, p < 0.01$), indicating that the performance premium of single-sex schooling is significantly smaller for girls than for boys. More importantly, the coefficient on *High-stakes X Single-sex* (α_6) is also

Table 4: Gender gap in response to increased stakes: the role of SSS

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
Female	-2.280*** (0.261)	-2.284*** (0.265)	-2.406*** (0.254)	-3.213*** (0.345)	-1.778*** (0.327)	-0.832** (0.368)
High-stakes exam		-0.129 (0.403)	-0.215 (0.407)	-0.946* (0.500)	-0.240 (0.535)	0.599 (0.498)
Single-sex school			5.367*** (0.941)	5.372*** (0.942)	7.351*** (1.121)	8.184*** (1.187)
Female X High-stakes exam				1.512*** (0.301)	1.595*** (0.297)	-0.208 (0.258)
Female X Single-sex					-2.641*** (0.615)	-4.337*** (0.785)
High-stakes X Single-sex					-1.337*** (0.351)	-2.866*** (0.482)
Female X High-stakes X Single-sex						3.188*** (0.533)
Constant	101.102*** (0.126)	101.172*** (0.294)	98.242*** (0.581)	98.640*** (0.610)	97.580*** (0.647)	97.131*** (0.665)
r2	0.02	0.02	0.04	0.04	0.04	0.04
N	363428	363428	363428	363428	363428	363428

All regressions include district-year fixed effects.

Standard errors in parentheses, clustered at district-year level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

negative and significant ($\alpha_6 = -1.337, p < 0.01$), suggesting that for boys, the advantage of attending a single-sex school diminishes in high-stakes exams.

The final specification in column (6) introduces the triple interaction term, *Female X High-stakes X Single-sex*, and presents the main finding of our paper. Note that since our dependent variable is a standardized score, all these effects are in *relative terms*. The results describe how each group's position in the national performance distribution shifts relative to others when moving from a low- to a high-stakes setting. The two-way interaction *Female X High-stakes exam* (α_3), which was positive and significant in all previous specifications, is now small and statistically insignificant ($\alpha_3 = -0.208$). This implies that in coeducational schools, there is no differential performance effect for girls under pressure. Instead, the entire effect is absorbed by the triple interaction term. The coefficient α_7 is large, positive, and highly significant ($\alpha_7 = 3.188, p < 0.01$). I.e., the enhanced performance of girls under pressure is driven entirely by those studying in single-sex schools.

A different way to interpret the results in column (6) of Table 4 is by analyzing how the gender gap in mathematics performance evolves under pressure, separately for students in coeducational and single-sex schools. First, let's consider students in coeducational schools. The gender gap in the low-stakes exam is given by the coefficient on *Female* (α_1), which is -0.832 points. This indicates a baseline performance gap in favor of boys. In the high-stakes exam, this gender gap becomes the sum of $\alpha_1 + \alpha_3$, which is $-0.832 - 0.208 = -1.040$ points. Therefore, in coeducational schools, the move to a high-stakes environment is associated with a slight *widening* of the gender gap by 0.21 points, further favoring boys. Although this change is not statistically significant, it aligns with the direction of findings in some of the existing literature where girls' performance declines relative to boys' under pressure.

The story is dramatically different for students in single-sex schools. In the low-stakes exam, the gender gap for students in single-sex schools is the sum of the baseline gap and the interaction with single-sex schooling: $\alpha_1 + \alpha_5 = -0.832 - 4.337 = -5.169$ points. This reveals a very large initial performance gap in favor of boys in single-sex schools. However, when moving to the high-stakes exam, this gap changes substantially. The gender gap in the high-stakes exam for this group is given by the sum of all relevant coefficients: $\alpha_1 + \alpha_3 + \alpha_5 + \alpha_7 = -0.832 - 0.208 - 4.337 + 3.188 = -2.189$ points.

Therefore, while the gender gap in coeducational schools widens slightly, the gap in single-sex schools *narrows* by 2.98 standardized points ($-2.189 - (-5.169)$). This is precisely the economic and statistical magnitude of the triple interaction coefficient, α_7 , net of the small change in coed schools.

In sum, the story that emerges is one of starkly different behavioral responses to pressure. While girls in coeducational schools do not manage to close the gap with boys, those in single-sex schools do so, and by a significant margin. This is all the more surprising given that girls in single-sex schools already started with an advantage compared to their peers in coeducational settings. A possible explanation is that the single-sex environment fosters a form of academic ambition in girls that encourages them to sustain high levels of effort, allowing them to close the gap with boys precisely when it matters most. In the next section, we use detailed survey data to directly test this hypothesis by examining students' self-reported effort, stress, and well-being.

6 Mechanisms

6.1 Ruling Out Observable School Characteristics as Drivers

To pinpoint whether any observable features that distinguish single-sex from co-educational schools can account for the outcome patterns we document, we examine three potential channels: (i) the share of female peers, (ii) the share of female teachers, and (iii) public versus private governance. Those three characteristics were particularly different between SS and Coed schools based on Table A.1. A naive approach would be to include these variables in our main specification; however, because single-sex schools sit at the extremes on all three dimensions (see Table A.1), this risks conflating the single-sex school effect with any one of these characteristics.

To avoid this, we adopt a more rigorous diagnostic approach. We restrict our analysis to the **coeducational subsample only**, where each of these three variables varies continuously. We then replace the *Single-sex* indicator in our DDD model with each of these characteristics in turn. If one of these variables is the true mechanism, it should generate the same significant triple interaction that we found for single-sex schooling. As we show below, none of them do.

6.1.1 The Role of Female Peers

First, we test whether the key factor is simply the presence of more female peers, rather than the complete absence of male peers. We re-run our DDD model on the coed

subsample, replacing the *Single-sex* dummy with a continuous variable for the share of female students in the school. The results are presented in Table 5.

Table 5: Gender gap in response to increased stakes: the role of Female Peers in COED schools

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
Female	-0.852** (0.343)	-0.848** (0.340)	-1.378*** (0.270)	-1.389*** (0.277)	4.906 (3.204)	5.328 (3.651)
High-stakes exam		0.215 (0.428)	0.117 (0.444)	0.107 (0.485)	4.049* (2.256)	4.488* (2.465)
Share of Female Students			26.606** (10.236)	26.607** (10.236)	37.484*** (9.707)	37.938*** (9.407)
Female X High-stakes exam				0.022 (0.236)	0.285 (0.198)	-0.593 (2.384)
Female X Share of Female Students					-13.742** (6.805)	-14.644* (7.738)
Share of Female Students X High-stakes exam					-8.704* (4.820)	-9.668* (5.258)
Female X High-stakes exam X Share of Female Students						1.876 (5.097)
Constant	97.741*** (0.159)	97.625*** (0.227)	85.487*** (4.732)	85.492*** (4.732)	80.568*** (4.488)	80.362*** (4.353)
r ²	0.04	0.04	0.04	0.04	0.05	0.05
N	157887	157887	157887	157887	157887	157887

All regressions include district-year fixed effects.

Standard errors in parentheses, clustered at district-year level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The final specification in column (6) tells a clear story. The triple interaction term, *Female X High-stakes exam X Share of Female Students*, is small (1.876) and statistically insignificant. This indicates that merely increasing the proportion of female peers within a coeducational setting does not generate the dramatic narrowing of the gender gap under pressure that we observe in all-girls schools. The institutional context of a 100% female environment appears to be fundamentally different from that of a majority-female one.

Additionally, some of the two-way interactions suggest interesting patterns. The coefficient on *Female × Share of Female Students* is negative and statistically significant (-14.644), implying that in coeducational schools, girls actually perform worse relative to boys as the share of female peers rises. Similarly, the coefficient on *Share of Female Students × High-stakes exam* is negative (-9.668), suggesting that male students may also underperform under pressure when surrounded by more female peers. Together, these results hint that a higher concentration of female peers in coeducational environments might reduce overall relative performance under pressure for both genders, in sharp contrast to the positive effect observed in single-sex schools. This strengthens our interpretation that the institutional setting of an all-girls school operates through mechanisms distinct from the mere proportion of female peers.

6.1.2 The Role of Female Teachers

Next, we investigate the role of teacher gender. It is plausible that a higher share of female teachers could serve as role models, influencing student behavior. We test this by running the DDD model on the coed subsample, this time interacting our variables with the share of female teachers. The results are in Table 6.

Table 6: Gender gap in response to increased stakes: the role of Female Teachers in Coed schools

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
Female	-0.852** (0.343)	-0.848** (0.340)	-0.836** (0.341)	-0.786** (0.375)	2.122* (1.195)	1.878 (1.428)
High-stakes exam		0.215 (0.428)	0.217 (0.428)	0.261 (0.474)	-1.068 (1.053)	-1.274 (1.283)
Share of Female Teachers			-1.431 (2.202)	-1.432 (2.202)	-0.478 (2.827)	-0.659 (2.931)
Female X High-stakes exam				-0.095 (0.245)	-0.078 (0.237)	0.392 (0.910)
Female X Share of Female Teachers					-4.769*** (1.661)	-4.367** (2.046)
Share of Female Teachers X High-stakes exam					2.156 (1.479)	2.497 (1.781)
Female X High-stakes exam X Share of Female Teachers						-0.769 (1.248)
Constant	97.741*** (0.159)	97.625*** (0.227)	98.493*** (1.372)	98.469*** (1.401)	97.894*** (1.806)	98.003*** (1.880)
r ²	0.04	0.04	0.04	0.04	0.04	0.04
N	157887	157887	157887	157887	157887	157887

All regressions include district-year fixed effects.

Standard errors in parentheses, clustered at district-year level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Again, the evidence in column (6) points away from this being the primary channel. The triple interaction, *Female X High-stakes exam X Share of Female Teachers*, is small (0.769) and not statistically significant. This finding suggests that while teacher gender may be important for other outcomes, as documented by [Lim and Meer \(2020\)](#) in the Korean context, it does not explain the specific dynamic of the gender gap’s response to high-stakes pressure.

Looking more closely at column (6), the two-way interaction between *Female* $\times$ *Share of Female Teachers* is negative and statistically significant (-4.367). This indicates that in coeducational schools, girls actually perform worse relative to boys as the proportion of female teachers rises. At the same time, the coefficient on *Share of Female Teachers* $\times$ *High-stakes exam* is positive (2.497), although not statistically significant, hinting that male students might perform slightly better under pressure when more female teachers are present. Taken together, these results reinforce that the gender composition of teachers does not replicate the sharp narrowing of the gender gap observed in single-sex schools. Instead, the evidence points toward the institutional context of a fully female student body—rather than teacher gender—as the critical mechanism driving the observed behavioral differences under pressure.

6.1.3 The Role of Private School Status

Finally, we consider school ownership. Since most single-sex schools in our sample are private, one might argue that our main finding is driven by the autonomy, resources, or specific culture of private institutions rather than single-sex education itself. Previous work by [Hahn et al. \(2018\)](#) shows that private school students in Seoul outperform their public school peers, while [Kim \(2018\)](#) finds that privatization can alter school inputs like instructional time and teacher experience. To ensure our single-sex effect is not simply a proxy for these private school characteristics, we test this channel directly. We restrict

the analysis to coed schools and use a dummy for *Private school* status in our DDD specification. The results are shown in Table 7.

Table 7: Gender gap in response to increased stakes: the role of Private schools in COED schools

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
Female	-0.852** (0.343)	-0.848** (0.340)	-0.842** (0.341)	-0.792** (0.376)	-1.158*** (0.343)	-1.116*** (0.354)
High-stakes exam		0.215 (0.428)	0.217 (0.428)	0.261 (0.474)	0.298 (0.481)	0.336 (0.462)
Private school			0.477 (0.968)	0.477 (0.968)	-0.179 (1.181)	-0.091 (1.240)
Female X High-stakes exam				-0.095 (0.246)	-0.087 (0.244)	-0.167 (0.226)
Female X Private					1.670* (0.939)	1.480 (1.114)
Private X High-stakes exam					-0.198 (0.558)	-0.365 (0.743)
Female X High-stakes exam X Private						0.366 (0.669)
Constant	97.741*** (0.159)	97.625*** (0.227)	97.517*** (0.301)	97.493*** (0.315)	97.643*** (0.317)	97.623*** (0.314)
r2	0.04	0.04	0.04	0.04	0.04	0.04
N	157887	157887	157887	157887	157887	157887

All regressions include district-year fixed effects.

Standard errors in parentheses, clustered at district-year level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results in column (6) once again show no evidence that private status can replicate the single-sex effect. The triple interaction term, *Female X High-stakes exam X Private*, is small (0.366) and far from statistical significance. This rules out the possibility that our main result is simply a proxy for the well-documented performance effects of private schooling.

6.1.4 School Level Analysis: Performance Distributions

In this section we turn to an examination of the performance distributions at the school level. If the effects of single-sex schooling are indeed driven by a behavioral response from students rather than by school-level inputs, as we have shown before, we should see evidence of this in how the entire distribution of performance evolves under pressure. To investigate this, Figure 1 provides a non-parametric visualization of the school-level average score distributions. The plots show the density of *school-level* average scores and are based on raw, unadjusted data. This approach allows us to move beyond the average treatment effects estimated in our regression models and observe whether the narrowing of the gender gap is due to a uniform shift or a change in the shape of the distributions.

Moving from the bottom panel to the upper one, and comparing the distributions of all-boys and all-girls single-sex schools, one can see how the gender gap in performance varies with high stakes. The mathematics plots (left column) visually corroborate the key findings from our DDD model in Table 4. The transition from the low-stakes

(bottom-left) to the high-stakes (top-left) exam reveals two distinct changes. First, the distribution for all-girls schools shifts rightward. Both the left and right tails of the distribution move to a higher performance level, suggesting a broad-based improvement that affects girls' schools across the entire performance spectrum. Second, the distribution for all-boys schools undergoes a bigger and similarly symmetric change but in the opposite direction. As a result of these changes, the gender gap in performance decreases significantly in the high stakes exam.

The fact that the entire distribution for all-girls schools shifts positively, while the entire distribution for all-boys schools shifts negatively, suggests a general and pervasive phenomenon. This is not a story about only elite schools or low-performing schools reacting differently; rather, it indicates a behavioral response that is common to single-sex schools regardless of their specific characteristics. This finding is consistent with our analysis in the previous sections, which showed that observable school-level factors cannot explain our results.

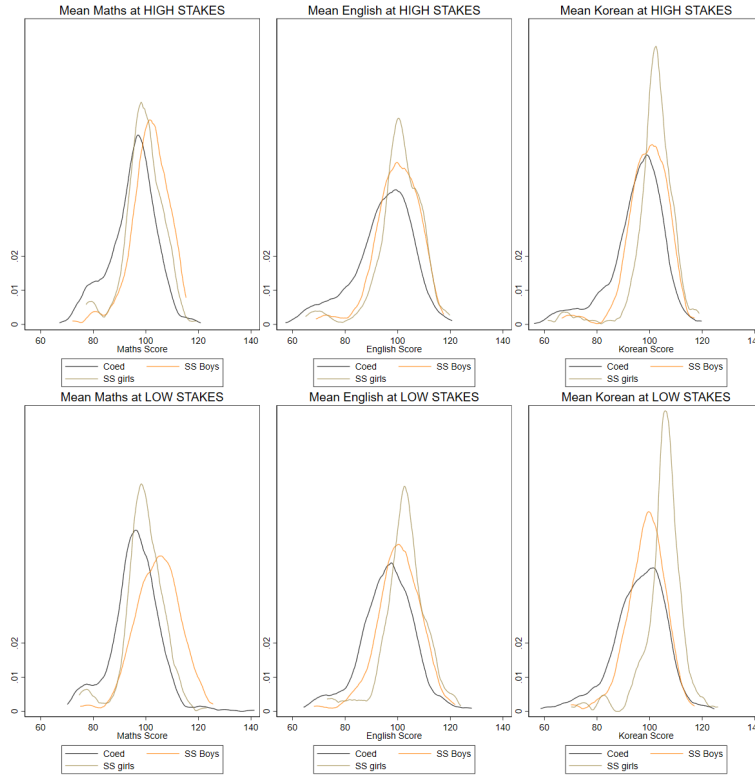


Figure 1: Kernel Density of School-Level Average Scores by School Type and Stakes

Notes: The figure plots the kernel density of school-level average scores for three school types: coeducational (black), single-sex boys (orange), and single-sex girls (green). The top row shows distributions for the high-stakes CSAT exam, and the bottom row shows distributions for the low-stakes NAEA exam. The x-axis represents the standardized score, and the y-axis represents the density.

7 Evidence on Behavioral Mechanisms from Survey Data

While our analysis of administrative data reveals *what* happens—that the gender gap in math narrows under pressure only in single-sex schools—it cannot fully explain *why*. To explore the underlying behavioral mechanisms, we turn to the Korean Youth Panel Survey (KYPS). This longitudinal dataset provides detailed information that is unavailable in the administrative records. Specifically, it allows us to track students’ self-reported stress levels, their self-assessed academic performance, and, crucially, their study habits over time. We can observe how many hours they dedicate to studying alone and their participation in private tutoring—a key component of the Korean educational landscape. By examining how these behavioral and psychological variables evolve from middle school through the high-stakes exam period and by type of school, we can test whether the single-sex school environment is associated with differential changes in effort and well-being that could explain the performance patterns we have documented.

7.1 How does SS schooling impact students’ stress, performance, and effort over time?

We estimate the following event study model for a range of outcomes related to stress, self-assessed performance, and effort:

$$Y_{it} = \sum_{\tau \neq -1} \beta_{\tau} \cdot \mathbb{I}\{\text{Time}_{it} = \tau\} \times \text{SS}_i + \mu_i + \varepsilon_{it} \quad (4)$$

Here, the coefficient β_{τ} captures the differential outcome for students in single-sex schools at event time τ , relative to their peers in coeducational schools and relative to the baseline period before high school entry ($\tau = -1$). We trace these effects from middle school ($\tau < 0$) through high school ($\tau = 0, 1, 2$) and, depending on the outcome, into university ($\tau = 3$). All individuals in the sample are in coed school during middle school, hence the event indicates exposure to single-sex schooling for the first time. The outcome variables are described in Table [8](#).

Table 8: Descriptive statistics: Outcomes of the Event Study

	Mean	StdDev.	N
<i>School stress</i>			
I get stressed by my parents' concern on my sch. grades (1 to 5)	2.90	(1.05)	3436
I get stressed by poor sch. grades (1 to 5)	3.08	(1.09)	3434
I get stressed by home assignments or examinations (1 to 5)	3.19	(1.08)	3434
<i>Performance last semester</i>			
Self-assessed performance in Korean (1 to 5)	3.27	(0.82)	3000
Self-assessed performance in English (1 to 5)	3.04	(1.03)	3001
Self-assessed performance in Math (1 to 5)	3.04	(1.12)	2999
<i>Mental Health</i>			
I am not interested in anything	2.22	(0.85)	2807
Sometimes I feel extremely anxious with no apparent reason	2.75	(1.10)	2803
Sometimes I feel suicidal with no apparent reason	2.05	(1.05)	2806
<i>Effort</i>			
Time spent studying by oneself : Korean (hours/week)	2.22	(2.74)	2980
Time spent studying by oneself : English (hours/week)	2.91	(3.36)	2980
Time spent studying by oneself : Maths (hours/week)	3.09	(3.76)	2980
Private tutoring-Korean (dummy)	0.45	(0.50)	3004
Private tutoring-English (dummy)	0.66	(0.47)	3004
Private tutoring-Maths (dummy)	0.69	(0.46)	3004
Time spent on private tutoring: Korean (hours/week)	3.39	(1.85)	1342
Time spent on private tutoring: English (hours/week)	4.12	(2.25)	1967
Time spent on private tutoring: Maths (hours/week)	4.28	(2.45)	2039

7.1.1 The Impact on Stress and Self-Perception

We first examine the psychological impact of single-sex schooling measured by the level of stress and self-perception. Figure 2 plots the effect on self-reported stress across three dimensions: stress induced by parents (*fam_grades*), stress from self-perceived low school grades (*sch_stress1*), and stress from homework assignments and exams (*sch_stress2*) (see, Table 8). The top row (for girls, in pink) shows that upon entering single-sex high school ($\tau = 0$), girls report a significant increase in stress, both from self-perceived low scores (*sch_stress1*) and from homeworks and exams (*sch_stress2*). During the first year of high school ($\tau = 1$), the effect is an increase of approximately 0.5 and 0.25 points, respectively, on a 5-point scale. For boys (bottom row, blue), there is no such pattern, with all point estimates small and statistically insignificant.

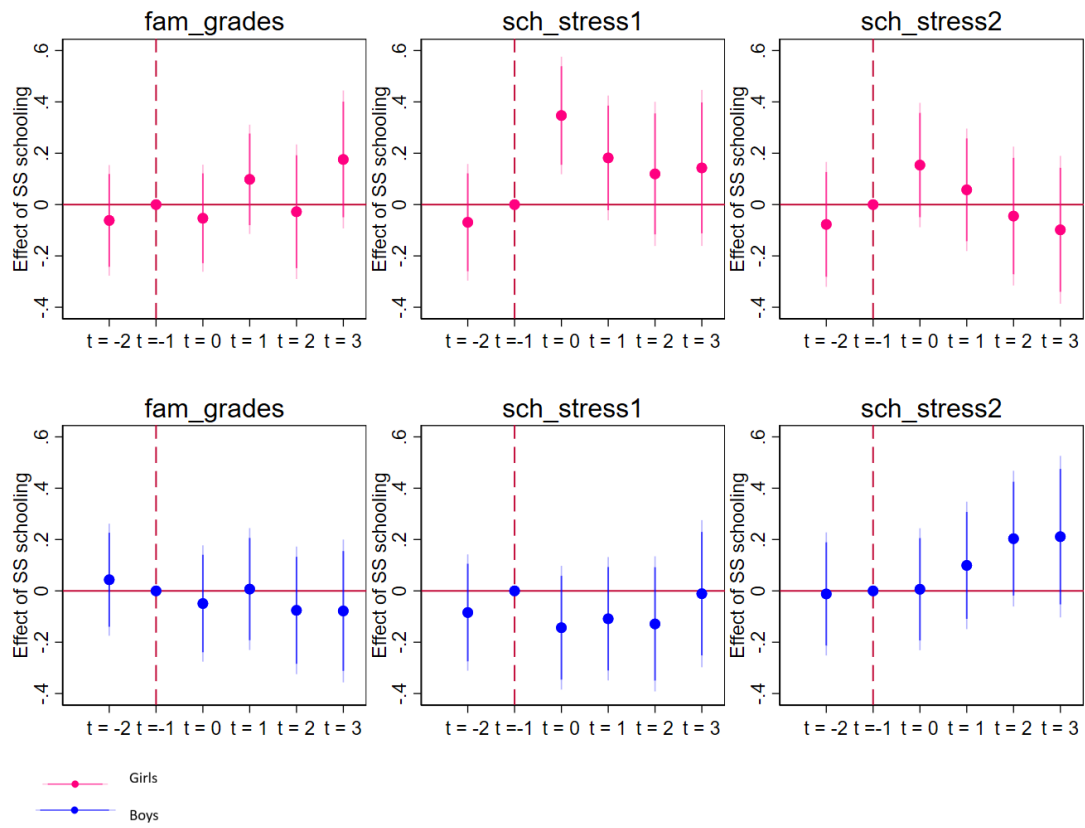


Figure 2: Event study: Effect of SS schooling on Stress from school over time.

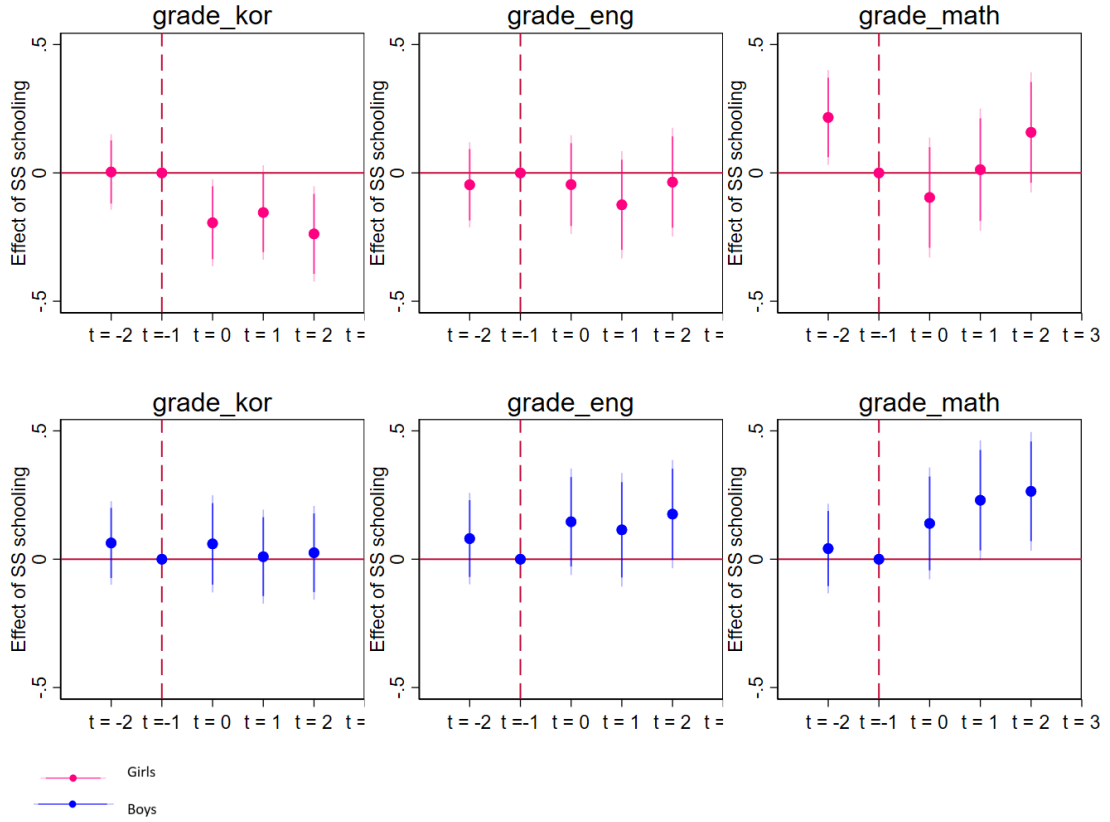


Figure 3: Event study: Effect of SS schooling on Self-Assessed Performance by Subject from school over time.

This psychological pressure is also reflected in students' self-perceptions, as shown in Figure 3. The results reveal a sharp disconnect between self-assessed performance and the objective reality documented in our administrative data. For girls, attending a single-sex school is associated with an immediate drop of about 0.25 points in their self-assessed math performance (*grade_math*), statistically significant. This increase in self-criticism occurs precisely as their relative performance holds strong. Conversely, boys in single-sex schools exhibit no such drop in self-assessed math performance; if anything, their self-perception trends slightly upward over time. This happens even as their objective exam scores show a significant relative decline under high-stakes pressure. Together, these patterns suggest that girls in single-sex schools internalize the pressure by becoming harsher self-critics, while boys may become more disconnected from their actual performance.

7.1.2 Broader Mental Health Costs of Single-Sex Schooling

The psychological costs associated with single-sex schooling for girls appear to extend beyond academic-related stress into broader measures of mental well-being. Figure 4 examines the effect of single-sex schooling on three indicators of psychological distress: feelings of disinterest (*feel_interest*), general anxiety (*feel_anxious*), and suicidal ideation (*feel_suicidal*). The results reveal a clear and concerning divergence between girls and boys.

For girls (top row), attending a single-sex school is associated with a marked decline in mental well-being. The point estimates for all three outcomes trend upwards after they enter high school at $\tau = 0$. Specifically, girls in single-sex schools report an increase in feelings of disinterest, anxiety, and, most alarmingly, suicidal ideation compared to their peers in coeducational schools. While the confidence intervals are wide, the consistency of the positive coefficients across all three measures points to a troubling pattern. The effect on suicidal ideation, for instance, appears to peak at $\tau = 2$, during the intense final year of preparation for the CSAT exam.

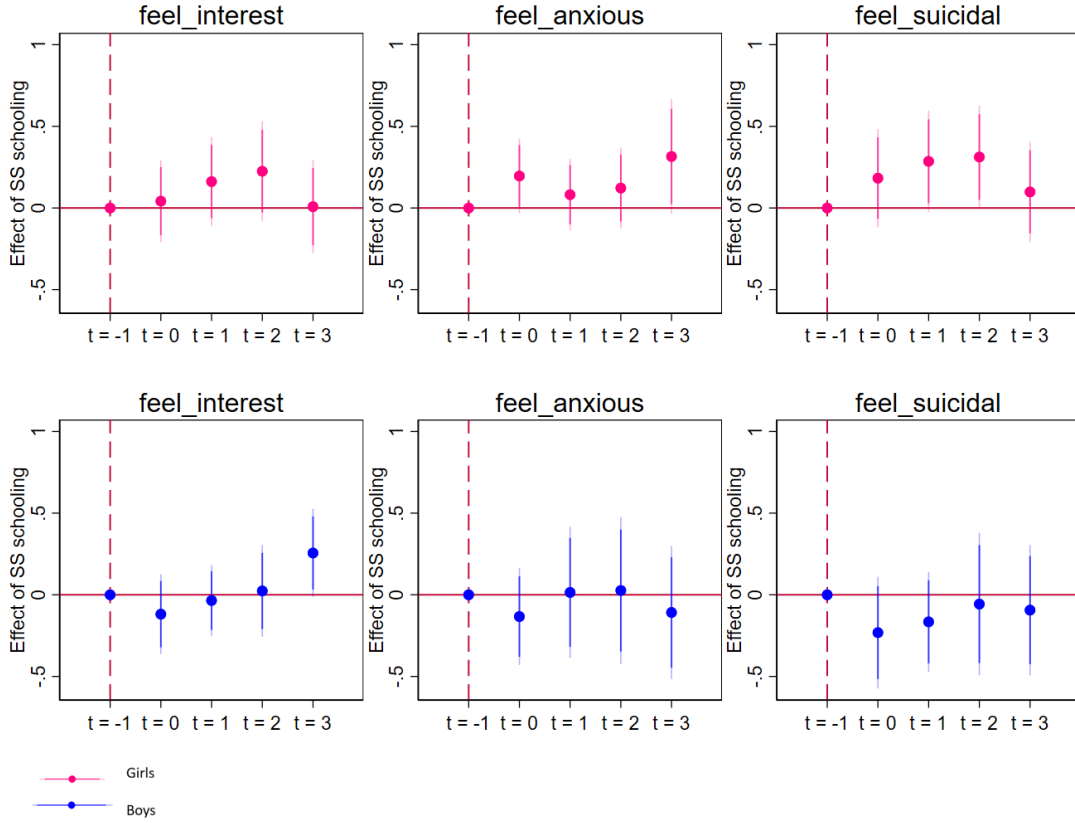


Figure 4: Event Study: Effect of SS Schooling on General Mental Well-being

In stark contrast, the results for boys (bottom row) show no evidence of a similar negative psychological impact. The coefficients for all three outcomes are clustered around zero and are not statistically significant, with the only exception of feelings of disinterest during the first year in university. In the case of suicidal ideation, the point estimates are consistently negative, suggesting that if anything, boys in single-sex schools may experience slightly better mental health outcomes than their coeducational peers, though this effect is not statistically significant.

7.1.3 The Impact on Effort: Study Time and Private Tutoring

The increased stress and critical self-assessment among girls in single-sex schools are directly linked to a substantial increase in effort. Figure 5 shows the effect on hours spent studying alone in each of the three subjects. For girls, attending a single-sex school leads to a steady and significant increase in study time for Math, peaking at an additional 2 hours per week in the final year of high school ($\tau = 2$). No such increase is observed for English or Korean, nor is there any significant effect for boys in any subject.

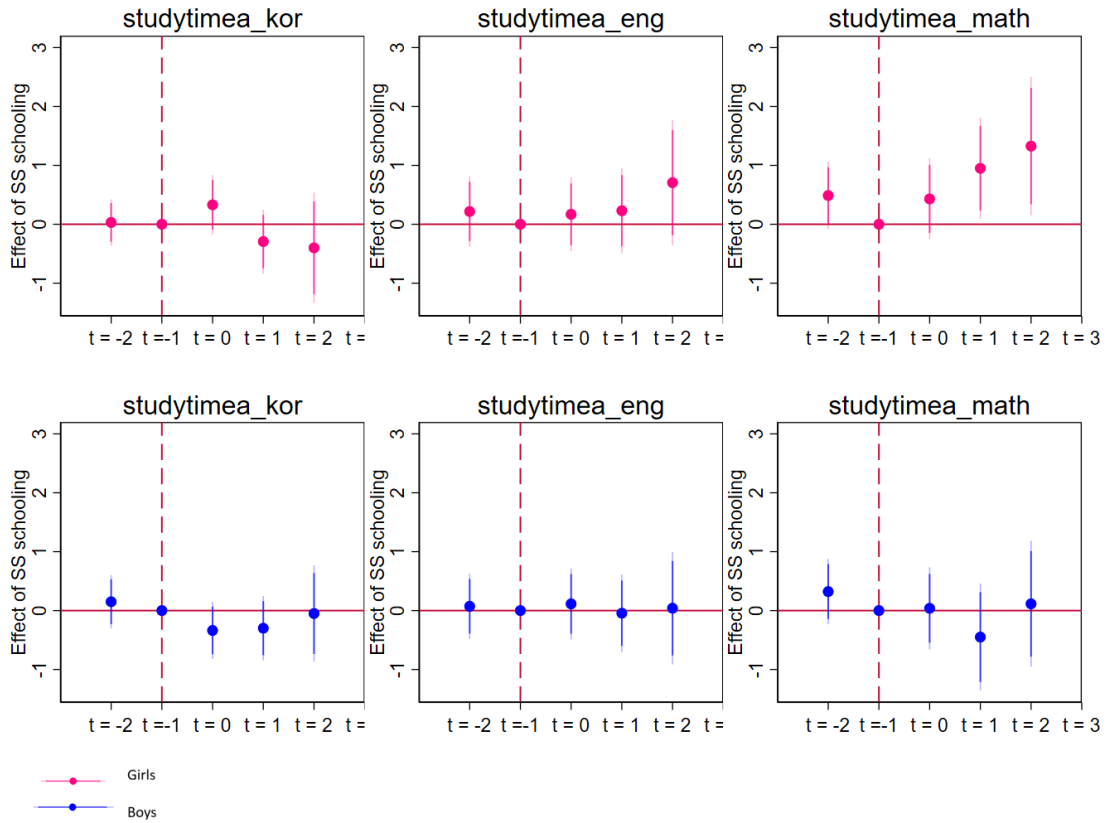


Figure 5: Event study: Effect of SS schooling on Time spent studying alone over time.

This pattern of intensified effort is even starker when we look at private tutoring. Figure 6 shows the extensive margin (whether a student receives any tutoring), and Figure 7 shows the intensive margin (hours of tutoring). For girls in math, single-sex schooling leads to a roughly 10 percentage point increase in the likelihood of receiving private tutoring in the first year of high school, statistically significant, which raises to a 20 percentage points increase by the final high school year. This is in contrast to a statistically significant decrease of between 10 to 15 percentage points for boys in English and Math tutoring. In the case of the intensive margin, tutoring time, the effects are in general noisy and non-statistically significant (Figure 7).

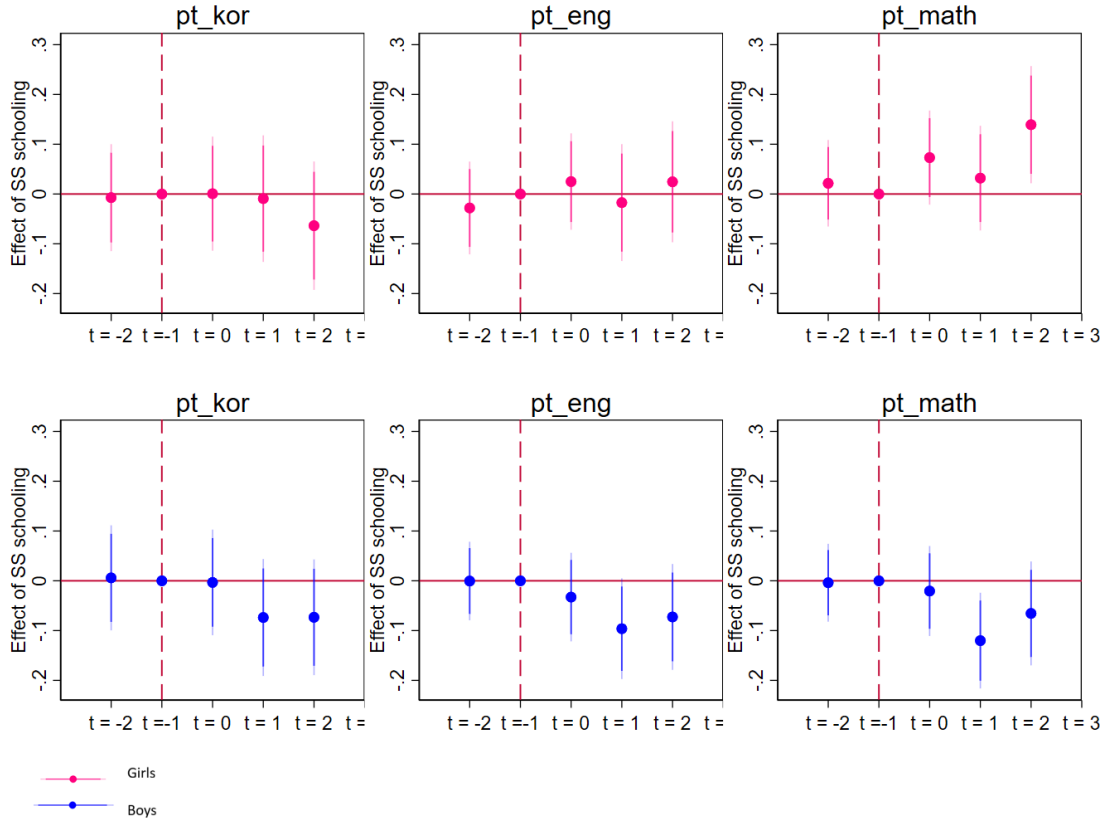


Figure 6: Event study: Effect of SS schooling on Effort: Private tutoring (Extensive margin 0/1) over time.

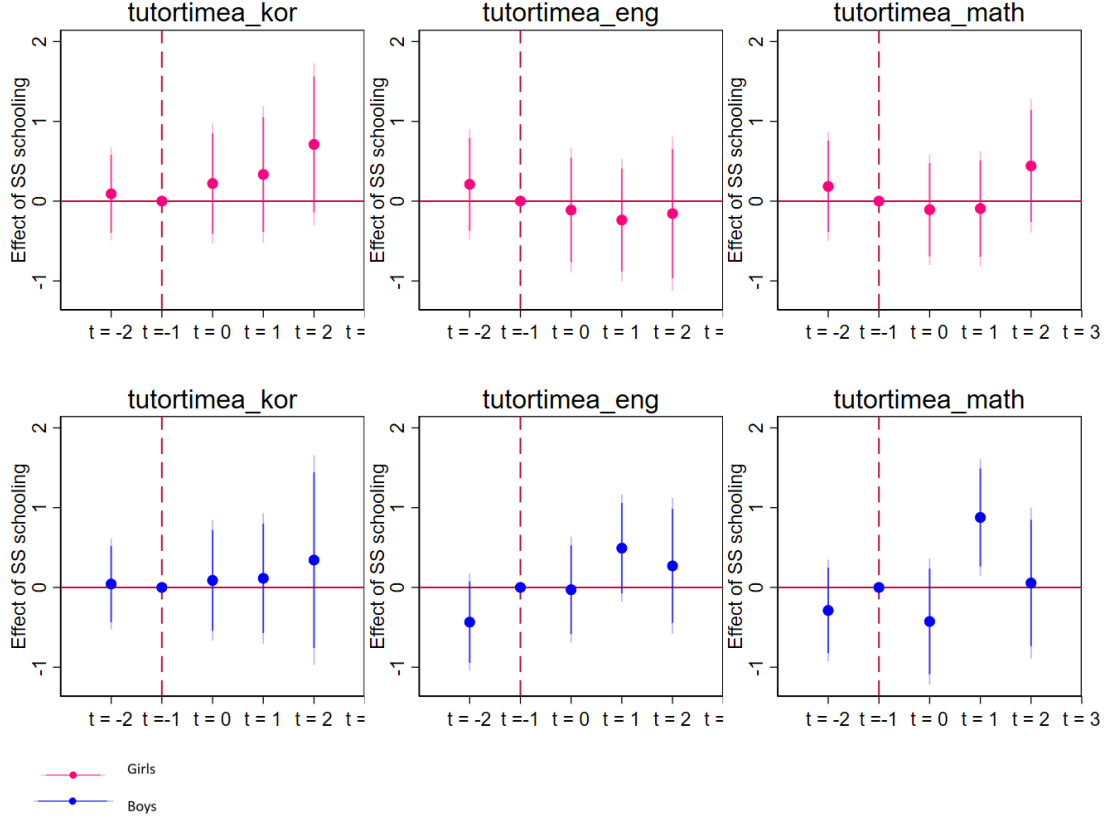


Figure 7: Event study: Effect of SS schooling on Effort: Private tutoring (Intensive margin (hours)) over time.

In sum, the event study analysis shows a clear behavioral contrast between boys and girls that complements our findings from the administrative data. Freed from the social dynamics of a coeducational environment, girls respond to the increasing academic pressure of high school by channeling their energy into the most competitive subject: mathematics. They study more, get more tutoring, and, as a result, improve their relative performance in high-stakes exams. This comes at a significant cost to their well-being, reflected in higher stress and harsher self-perceptions. Boys in single-sex schools, in contrast, do not exhibit this behavioral shift, which helps explain why their relative performance advantage deteriorates when the stakes are highest.

8 Long Term Implications: Enrollment in STEM Majors

Beyond test scores, a crucial question is whether the behavioral responses of attending a single-sex school translates into different long-term educational choices for girls. To

explore this, we use KYPs data on the actual university major each student enrolled in after graduation. We estimate a linear probability model to see how single-sex schooling affects the likelihood that a student enrolls in a STEM field. The results are presented in Table 9.

The columns in the table represent different definitions of STEM, moving from the narrowest and most quantitatively-intensive fields in column (1) to broader definitions in subsequent columns. In column (4), the definition of STEM includes all scientific fields *except biology*. This classification is particularly insightful, as biology is a field that tends to attract more women and can mask gender gaps in more male-dominated areas like engineering, computer science, and physics. Across all specifications, the coefficient on *Female* is large, negative, and highly significant. This confirms the well-documented pattern that girls are, on average, 25-30 percentage points less likely than boys to enroll in a STEM major. The main effect of single-sex schooling, given by the coefficient on *Single-sex school*, is small and not statistically significant across all models, indicating that boys' choice of university major is not affected by the type of high school.

The interaction term, *Female X Single-sex*, captures whether the gender gap in STEM enrollment is different for girls who attended single-sex schools. A clear pattern emerges as we adjust the definition of STEM. The point estimate of the interaction term is consistently positive and grows in magnitude as the definition of STEM becomes more focused on traditionally male-dominated fields. The coefficient increases from 0.058 in the broad definition (column 2) to 0.071 in the narrower definition (column 1) to 0.106 in column (4), where the definition of STEM includes all scientific fields *except biology*. This last coefficient estimate is marginally significant with a *p-value* of 0.115. While these effects are not statistically significant, they are economically large. The magnitude of the coefficient in column (4) implies that attending a single-sex school is associated with a 10.6 percentage point increase in the probability that a girl will enroll in a non-biology STEM major—an effect large enough to close over one-third of the baseline gender gap for that outcome (-0.306).

All in all, although we cannot draw a definitive conclusion due to the lack of statistical precision, the consistent direction and increasing magnitude of the effect across specifications provide suggestive evidence that the single-sex school environment may ultimately lead girls to enroll in the most quantitatively-intensive and traditionally male-dominated fields.

9 Conclusions

This paper investigates how the school peer environment shapes gender differences in performance under pressure. Leveraging a lottery-based assignment system in South Korea, we compare student outcomes in low-stakes and high-stakes national exams for students in single-sex versus coeducational schools. Our findings reveal a sharp divergence in responses. Contrary to much of the literature, we find that the gender gap in math performance narrows significantly in high-stakes exams. However, this effect is driven entirely by students in single-sex schools.

Table 9: The Effect of Single-Sex Schooling on STEM Major Choice

	(1) STEM_narrowest	(2) STEM_narrow	(3) STEM_wide	(4) STEM_nobio
Female	-0.251*** (0.043)	-0.288*** (0.060)	-0.293*** (0.062)	-0.306*** (0.061)
Single-sex school	-0.011 (0.052)	-0.022 (0.054)	-0.031 (0.055)	-0.072 (0.053)
Female X Single-sex	0.071 (0.052)	0.058 (0.063)	0.066 (0.063)	0.106 (0.066)
R2	0.11	0.13	0.13	0.13
N	1088	1088	1088	1088

Standard errors in parentheses, clustered at level of FE (school districts).

Controls include:

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Our analysis with survey data indicates that girls in single-sex schools are more strict with themselves, ramp up their study effort and private tutoring as exams approach, though this comes at the cost of higher reported stress. Our results are not explained by differences in teacher gender, peer gender ratios within co-ed schools, or private school status.

These findings contribute to the *nature versus nurture* debate by showing that a key environmental factor—the absence of the other gender—can fundamentally alter behavior in high stakes situations. They support theories of social norms, suggesting that girls in coeducational settings may feel pressure to downplay academic ambition, a pressure that is alleviated in an all-girls environment.

Appendix: Additional tables and figures

Table A.1: School characteristics: Single-sex versus coed.

Variable	SS boys	SS girls	Coed	All
Share of female students	0.00***	1.00***	0.46	0.49
Share of female teachers	28.68***	52.84***	59.04	49.37
Number of teachers	73.27	73.94	74.15	73.86
Students in 1st year	399.97	398.36	393.99	396.80
Students in 2nd year	401.14	395.70	385.81	392.65
Students in 3rd year	407.60*	399.89	386.09	395.65
Pupils per teacher	16.40***	16.08***	15.46	15.88
Year of creation	1964.87***	1966.49***	1988.33	1975.97
Private school	0.72***	0.68***	0.24	0.49
Length school year: 1st	206.24	206.38	206.36	206.33
Length school year: 2nd	206.35	206.54	206.43	206.44
Length school year: 3rd	206.45	206.69	206.71	206.64
Share unionized teachers	43.85*	42.87	40.74	42.16
Students with scholarship	452.60*	487.00***	410.14	443.03
Students with reduced tuition	441.94	482.98***	406.43	437.40
Libraries	33.80	33.40	32.87	33.26
AV rooms	1.10	1.12*	1.03	1.07
Lunchrooms	0.59	0.64	0.67	0.64
School building area	2.69**	2.29	2.24	2.37
Library seats	86899.61	74790.21	20198.87	53322.63
Library seats	97.20*	89.13	79.29	86.79
Observations	163	174	280	617

Stars indicate statistical significance of Welch two-sample t -tests comparing single-sex and coeducational schools.

"SS boys" schools are compared to coed schools, and "SS girls" schools are compared to coed schools.

The unit of analysis in this table is a school.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: Distribution of Share of Females across Schools (Edudata 2011)

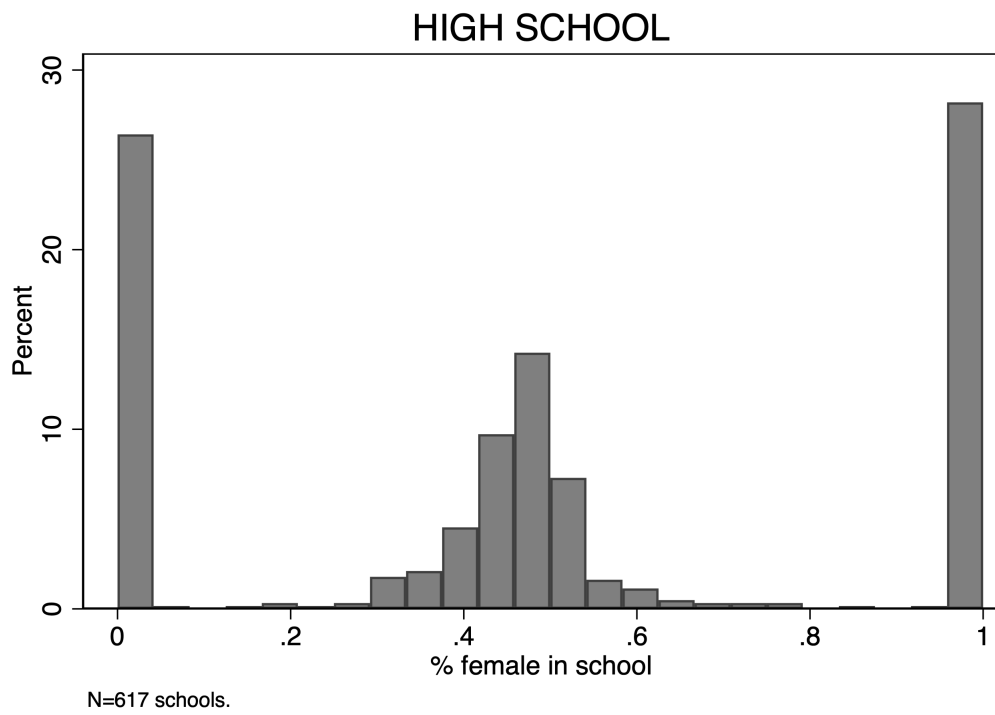


Figure A.2: Distribution of Female and Male Performance at CSAT and NAEA

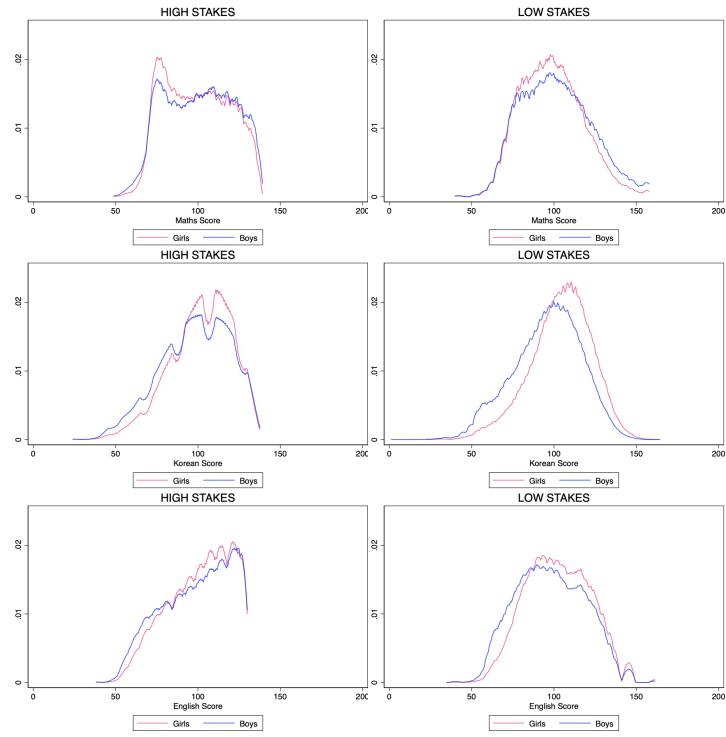


Table A.2: The role of SSS across subjects

	Maths			Korean	English
	(1)	(2)	(3)	(4)	(5)
Female	-0.832** (0.368)	-1.480*** (0.200)	0.000 (.)	0.000 (.)	0.000 (.)
High-stakes exam	0.599 (0.498)	-0.683*** (0.254)	0.714*** (0.000)	3.620*** (0.000)	2.114 (.)
Single-sex school	8.184*** (1.187)	0.000 (.)	7.646*** (0.000)	7.087*** (0.000)	6.908 (.)
Female X High-stakes exam	-0.208 (0.258)	0.170 (0.197)	-0.295 (0.410)	-5.862*** (0.434)	-3.723*** (0.327)
Female X Single-sex	-4.337*** (0.785)	0.000 (.)	-4.047*** (0.949)	-3.446*** (0.870)	-3.694*** (1.086)
High-stakes X Single-sex	-2.866*** (0.482)	-1.966*** (0.424)	-2.924*** (0.000)	-2.108*** (0.000)	-1.880 (.)
Female X High-stakes X Single-sex	3.188*** (0.533)	2.484*** (0.462)	3.175*** (0.392)	1.631*** (0.495)	2.346*** (0.350)
Constant	97.131*** (0.665)	101.275*** (0.131)	96.934*** (0.272)	96.931*** (0.241)	97.188*** (0.305)
FE	Distr-Year	Sch-Year	Distr-Year-Female	Distr-Year-Female	Distr-Year-Female
R2	0.04	0.14	0.03	0.05	0.04
N	363428	363428	363428	373022	370934

Dep. var. is score at either CSAT (HS) or NAEA (LS) exam.

Standard errors in parentheses, clustered at level of FE.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: KYPS-Descriptive statistics: Single-sex and coeducational schools-HIGH SCHOOL (“Old” cohort)

		All boys	All girls	Coed.	Total
2003 (2nd Middle)	# students	558	730	1250	2538
	% students	21.98	28.76	49.25	100
	# schools	18	24	41	83
2004 (3rd Middle)	# students	477	668	1135	2280
	% students	20.92	29.3	49.78	100
	# schools	24	29	42	95
2005 (1st High)	# students	630	744	494	1868
	% students	33.72	39.83	26.44	100
	# schools	157	168	100	425
2006 (2nd High)	# students	618	743	498	1859
	% students	33.24	39.97	26.79	100
	# schools	155	172	99	426
2007 (3rd High)	# students	603	695	471	1769
	% students	34.09	39.28	26.62	100
	# schools	155	171	101	427
2008 (University)	# students	8	106	1243	1357
	% students	0.59	7.81	91.6	100
	# schools	2	20	155	177
Total	# students	2894	3686	5091	11671
	% students	24.8	31.58	43.62	100
	# schools	511	584	538	1633

Note: These results were obtained by restricting to the sample of schools within the Equalization Policy area, where we observe at least 6 individuals of the same sex in a school the corresponding year, in order to infer whether a school is SS or coed.

Table A.4: Characteristics of **Girls** in Single-Sex Schools Vs Coed. schools- Year of random assignment (before entering high school)- NO District FE.

Variable	(1) Coed. schools	(2) (sd)	(3) SS schools	(4) (sd)	(5) Difference	(6) (se)
Log hh income	5.70	(0.52)	5.59	(0.48)	-0.11**	(0.04)
Father's education (years)	13.75	(2.68)	13.29	(2.82)	-0.46*	(0.26)
Mother's education (years)	12.61	(2.27)	12.32	(2.40)	-0.29	(0.25)
Father works FT	0.93	(0.26)	0.93	(0.26)	-0.00	(0.01)
Father works PT	0.04	(0.19)	0.04	(0.19)	-0.00	(0.02)
Mother works FT	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Mother works PT	0.08	(0.27)	0.07	(0.26)	-0.01	(0.02)
Age	14.79	(0.41)	14.80	(0.41)	0.01	(0.03)
born_early	0.19	(0.40)	0.18	(0.39)	-0.01	(0.03)
Has siblings	0.96	(0.21)	0.94	(0.24)	-0.02	(0.02)
Average performance	3.06	(0.80)	3.16	(0.72)	0.10*	(0.06)
Observations	223		720		943	

Standard errors in parentheses, clustered at industry level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Characteristics of **Girls** in Single-Sex Schools Vs Coed. schools- Year of random assignment (before entering high school)- With District FE.

Variable	(1) Coed. schools	(2) (sd)	(3) SS schools	(4) (sd)	(5) Difference	(6) (se)
Log hh income	5.70	(0.52)	5.59	(0.48)	-0.07	(0.05)
Father's education (years)	13.75	(2.68)	13.29	(2.82)	-0.24	(0.23)
Mother's education (years)	12.61	(2.27)	12.32	(2.40)	-0.06	(0.26)
Father works FT	0.93	(0.26)	0.93	(0.26)	-0.00	(0.02)
Father works PT	0.04	(0.19)	0.04	(0.19)	-0.00	(0.02)
Mother works FT	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Mother works PT	0.08	(0.27)	0.07	(0.26)	0.01	(0.02)
Age	14.79	(0.41)	14.80	(0.41)	0.03	(0.04)
born_early	0.19	(0.40)	0.18	(0.39)	-0.02	(0.03)
Has siblings	0.96	(0.21)	0.94	(0.24)	-0.02	(0.02)
Average performance	3.06	(0.80)	3.16	(0.72)	0.11	(0.08)
Observations	223		720		943	

Standard errors in parentheses, clustered at industry level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Characteristics of **Boys** in Single-Sex Schools Vs Coed. schools- Year of random assignment (before entering high school)- NO District FE.

Variable	(1) Coed. schools	(2) (sd)	(3) SS schools	(4) (sd)	(5) Difference	(6) (se)
Log hh income	5.69	(0.63)	5.54	(0.51)	-0.14**	(0.06)
Father's education (years)	14.09	(2.72)	13.47	(2.74)	-0.63**	(0.29)
Mother's education (years)	12.92	(2.33)	12.43	(2.39)	-0.48	(0.30)
Father works FT	0.95	(0.21)	0.92	(0.27)	-0.03	(0.02)
Father works PT	0.02	(0.15)	0.03	(0.18)	0.01	(0.02)
Mother works FT	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Mother works PT	0.09	(0.29)	0.06	(0.24)	-0.03	(0.02)
Age	14.77	(0.45)	14.79	(0.41)	0.02	(0.03)
born_early	0.21	(0.41)	0.20	(0.40)	-0.01	(0.03)
Has siblings	0.90	(0.30)	0.91	(0.29)	0.01	(0.03)
Average performance	3.28	(0.73)	3.28	(0.76)	0.00	(0.07)
Observations	235		599		834	

Standard errors in parentheses, clustered at industry level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Characteristics of **Boys** in Single-Sex Schools Vs Coed. schools- Year of random assignment (before entering high school)- With District FE.

Variable	(1) Coed. schools	(2) (sd)	(3) SS schools	(4) (sd)	(5) Difference	(6) (se)
Log hh income	5.69	(0.63)	5.54	(0.51)	-0.06	(0.09)
Father's education (years)	14.09	(2.72)	13.47	(2.74)	-0.14	(0.44)
Mother's education (years)	12.92	(2.33)	12.43	(2.39)	-0.24	(0.39)
Father works FT	0.95	(0.21)	0.92	(0.27)	-0.05*	(0.02)
Father works PT	0.02	(0.15)	0.03	(0.18)	0.01	(0.02)
Mother works FT	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Mother works PT	0.09	(0.29)	0.06	(0.24)	-0.01	(0.03)
Age	14.77	(0.45)	14.79	(0.41)	-0.01	(0.06)
born_early	0.21	(0.41)	0.20	(0.40)	0.00	(0.05)
Has siblings	0.90	(0.30)	0.91	(0.29)	-0.02	(0.03)
Average performance	3.28	(0.73)	3.28	(0.76)	-0.03	(0.08)
Observations	235		599		834	

Standard errors in parentheses, clustered at industry level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Gender gap in response to increased stakes: the role of Private schools

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
Female	-2.280*** (0.261)	-2.284*** (0.265)	-2.257*** (0.249)	-3.034*** (0.340)	-2.365*** (0.316)	-1.584*** (0.344)
High-stakes exam		-0.129 (0.403)	-0.126 (0.403)	-0.830 (0.497)	-0.524 (0.537)	0.195 (0.454)
Private school			2.343*** (0.779)	2.343*** (0.779)	3.373*** (0.817)	4.182*** (0.893)
Female X High-stakes exam				1.455*** (0.299)	1.449*** (0.296)	-0.016 (0.248)
Female X Private					-1.400** (0.623)	-3.028*** (0.849)
Private X High-stakes exam					-0.639* (0.364)	-2.112*** (0.542)
Female X High-stakes exam X Private						3.049*** (0.621)
Constant	101.102*** (0.126)	101.172*** (0.294)	100.031*** (0.463)	100.416*** (0.495)	99.920*** (0.456)	99.527*** (0.446)
r2	0.02	0.02	0.03	0.03	0.03	0.03
N	363428	363428	363428	363428	363428	363428

All regressions include district-year fixed effects.

Standard errors in parentheses, clustered at district-year level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

- Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, 95(4), 1438–1443.
- Azmat, G., Calsamiglia, C., & Iriberri, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6), 1372–1400.
- Booth, A., Fan, E., Meng, X., & Zhang, D. (2019). Gender differences in willingness to compete: The role of culture and institutions. *The Economic Journal*, 129(618), 734–764.
- Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542–555.
- Booth, A. L., Cardona-Sosa, L., & Nolen, P. (2018). Do single-sex classes affect academic achievement? an experiment in a coeducational university. *Journal of Public Economics*, 168, 109–126.
- Bursztyn, L., Egorov, G., & Jensen, R. (2019). Cool to be smart or smart to be cool? understanding peer pressure in education. *The Review of Economic Studies*, 86(4), 1487–1526.
- Buser, T., van den Assem, M. J., & van Dolder, D. (2023). Gender and willingness to compete for high stakes. *Journal of Economic Behavior & Organization*, 206, 350–370.
- Cai, X., Lu, Y., Pan, J., & Zhong, S. (2019). Gender gap under pressure: evidence from china’s national college entrance examination. *Review of Economics and Statistics*, 101(2), 249–263.
- Choi, E. J., Moon, H. R., & Ridder, G. (2014). Estimation of an education production function under random assignment with selection. *American Economic Review*, 104(5), 206–211.
- Doris, A., O’Neill, D., & Sweetman, O. (2013). Gender, single-sex schooling and maths achievement. *Economics of Education Review*, 35, 104–119.
- Dustmann, C., Ku, H., et al. (2018). Why are single-sex schools successful? *Labour Economics*, 54, 79–99.
- Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. W. (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in switzerland. *Journal of economic behavior & organization*, 115, 123–143.
- Fryer Jr, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210–240.
- Fryer Jr, R. G., & Torelli, P. (2010). An empirical analysis of ‘acting white’. *Journal of Public Economics*, 94(5-6), 380–396.
- Garcia-Brazales, J. (2025). Unlearning traditionalism: The long-run effects of schools on gender attitudes. *The Economic Journal*, ueaf070.
- Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5), 1637–1664.

- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3), 1049–1074.
- Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *The American Economic Review*, 94(2), 377–381.
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880), 1164–1165.
- Hahn, Y., & Wang, L. C. (2019a). The effectiveness of single-sex schools through out-of-school activities: Evidence from south korea. *Oxford Bulletin of Economics and Statistics*, 81(2), 369–393.
- Hahn, Y., & Wang, L. C. (2019b). The effectiveness of single-sex schools through out-of-school activities: Evidence from south korea. *Oxford Bulletin of Economics and Statistics*, 81(2), 369–393.
- Hahn, Y., Wang, L. C., & Yang, H.-S. (2018). Does greater school autonomy make a difference? evidence from a randomized natural experiment in south korea. *Journal of Public Economics*, 161, 15–30.
- Hara, H., & Rodriguez-Planas, N. (2025). Long-term consequences of teaching gender roles: Evidence from desegregating industrial arts and home economics in japan. *Journal of Labor Economics*, 43(2), 349–389.
- Iriberry, N., & Rey-Biel, P. (2019). Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *The Economic Journal*, 129(620), 1863–1893.
- Kearney, M. S., & Levine, P. B. (2025). *Why is fertility so low in high income countries?* (Tech. Rep.). National Bureau of Economic Research.
- Kim, Y. (2018). Privatization and school practices: Evidence from seoul’s high school choice policy. *International Journal of Educational Development*, 62, 322–332.
- Kleven, H. (2022). *The geography of child penalties and gender norms: A pseudo-event study approach* (Tech. Rep.). National Bureau of Economic Research.
- Kwak, D. W., Ku, H., et al. (2013). *Together or separate: Disentangling the effects of single-sex schooling from the effects of single-sex schools* (Tech. Rep.).
- Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2), 1–33.
- Lee, J., & Park, Y. (2017). Do single-sex school advantages exist? evidence from a school choice lottery program in seoul.
- Lee, S., Turner, L. J., Woo, S., & Kim, K. (2014). *All or nothing? the impact of school and classroom gender composition on effort and academic achievement* (Tech. Rep.). National Bureau of Economic Research.
- Lim, J., & Meer, J. (2020). Persistent effects of teacher–student gender matches. *Journal of Human Resources*, 55(3), 809–835.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3), 1067–1101.
- Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1), 601–630.

- Park, H., Behrman, J. R., & Choi, J. (2013). Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in seoul high schools. *Demography*, 50(2), 447–469.
- Schlosser, A., Neeman, Z., & Attali, Y. (2019). Differential performance in high versus low stakes tests: Evidence from the gre test. *The Economic Journal*, 129(623), 2916–2948.
- Wang, L. C. (2015). All work and no play? the effects of ability sorting on students? non-school inputs, time use, and grade anxiety. *Economics of Education Review*, 44, 29–41.