Response to Competition: Gender, Domains, and STEM choice *

Fanghua Li† Y. Jane Zhang‡

March 2024

Abstract

Women's lower performance in competitive environments has been advanced as an explanation for gender inequality in the labor market. We explore the domain specificity of gender differences in response to competition (RC) and test whether it can predict subsequent STEM track choice, an important contributor to the gender pay gap. Using Chinese administrative data, and defining RC as the performance improvement from a mock exam to the highly competitive High School Entrance Exam, we find a gender gap in RC favoring boys in STEM subjects, and favoring girls in non-STEM subjects. This pattern is persistent and individual RC is stable over time. Both domain-specific RC measures significantly predict subsequent STEM track choice, and contribute to 22% of the adjusted gender gap in STEM choice (controlling for STEM ability, comparative advantage in STEM, peer/role model effects, age, and SES). Principal component analysis reveals the presence of both a domain-general and a comparative advantage component of RC. Both components favor boys and both significantly predict STEM choice; only the comparative advantage component contributes to the gender gap in STEM choice.

*We thank the participants of the Gender, Norms and Economics Workshop 2022, the STEP UP Conference 2023, and seminar participants at the University of Technology Sydney, University of Melbourne, Deakin University, University of Auckland, Peking University, and Australian National University for their useful comments and discussions.

†University of New South Wales. Email: *fanghua.li@unsw.edu.au*

‡University of New South Wales. Email: *jane.zhang@unsw.edu.au*

1 Introduction

Competition is a cornerstone of modern society. How an individual responds to competition can influence every facet of their life. In particular, educational and labor market outcomes can depend heavily on performance in competitions, over and above one's abilities. A significant body of lab experimental studies find that men and women respond to competition differently. In their seminal paper, Gneezy et al. (2003) show that men's performance in a real-effort task improved when facing a competitive environment, but women's did not.[1] Relatedly, Niederle and Vesterlund (2007) find that women are inclined to avoid participating in competitions altogether even when there was no gender difference in performance in the underlying real effort task. Further research in this literature has found that women's lower willingness to compete in the lab can explain women's lower likelihood of choosing fields of study with high earnings potential such as STEM and math intensive tracks (Buser et al., 2014, 2017, 2024), lower salary expectations (Reuben et al., 2017), and lower starting salaries (Reuben et al., 2019). These findings suggest a different angle to understanding persistent gender inequality in the labour market: in addition to the traditional explanations of differential human capital accumulation and discrimination, the underrepresentation of women in high paying jobs and occupations could arise from the competitive nature of these positions (Bertrand, 2011; Niederle and Vesterlund, 2011; Shurchkov and Eckel, 2018).

In this paper we study gender differences in response to competition and the extent to which they predict a subsequent, irreversible decision to choose an academic track focusing on STEM. Gender differences in fields of study, and in STEM majors in particular, can account for significant portions of the gender wage gap among recent university

[1] Gneezy et al. (2003) conduct a lab experiment in which groups of 3 men and 3 women were given a maze-solving task and paid either on a non-competitive piece-rate scheme of $0.50 per maze or a competitive tournament scheme of $3 per maze but only the winner in the group of 6 is paid. They found that while men and women performed similarly well under non-competitive incentives, with competitive incentives men significantly improved their performance relative to women.

graduates (Francesconi and Parey, 2018; Card and Payne, 2021).[2] Existing studies largely consider the gender gap in response to competition as a domain-general phenomenon, with the potential to disadvantage women in any activity involving competition.[3] For instance, women's under-representation in STEM and in math-intensive fields is explained by the more competitive or prestigious nature of these fields (and that women tend to shy away from competition) (Niederle and Vesterlund, 2010; Buser et al., 2014, 2017, 2024).

We focus on domain-specific gender differences. In the context of a high-stakes educational competition, measuring response to competition as the performance improvement between the mock versus the real exam, we find a reversal of the gender gap across domains – while boys have higher response to competition in STEM subjects, girls have higher response to competition in non-STEM subjects. This pattern is persistent and individual response to competition is stable over time. Response to competition in both STEM and in non-STEM domains are significant predictors of subsequent STEM choice, together explaining 17% (22%) of the raw (adjusted) gender gap in choosing the STEM track.

Investigating the impact of domain-specific response to competition on STEM choice parallels recent findings that, in addition to STEM ability, comparative advantage in STEM over non-STEM subjects matters for the decision to study STEM (Valla and Ceci, 2014; Breda and Napp, 2019; Loyalka et al., 2017; Card and Payne, 2021; Goulas et al., 2022). Similarly, we find that comparative advantage in response to competition in STEM over non-STEM subjects is a significant determinant of STEM choice.

This paper uses administrative data from a county in Henan province, located in central

[2] More generally, there is substantial heterogeneity in earnings between college majors, and the earnings gap across majors can approach the overall wage premium of having a college degree (e.g., (Altonji et al., 2012)).

[3] Notable exceptions are Günther et al. (2010), Shurchkov (2012), and Cotton et al. (2013), discussed below. Cassar et al. (2016); Cassar and Rigdon (2021a,b); Cassar and Zhang (2022); Cassar and Rigdon (2023) show that women can be as competitive as men when the goals of competition are consistent with evolutionarily beneficial strategies for women.

China. The main analysis uses the universe of observations from four cohorts of students, each with approximately 5,000 students. Analogous to Gneezy et al. (2003), we observe individual performance under two environments, one competitive and one less so, for the same task. The competitive environment is an entrance exam for high school admissions in China, in which approximately 50% of students fail. The less competitive environment is a mock exam taken by the same students, administered by the same educational authorities, and follows the exact same format as the High School Entrance Exam (henceforth, HSEE). One cannot fail the mock exam but students have an incentive to perform to the best of their ability due to the informational value of mock exam scores for the high school admissions process. Empirically, the scores on the mock and real exam have correlation coefficients of around 0.95. We define response to competition as the difference in performance between the real and the mock exam at the individual level. For subsets of the cohorts, we additionally obtain the universe of observations on high school track choice (STEM or non-STEM) and college entrance exam (CEE) outcomes, which we link with individual response to competition. This allows us to study how the STEM choice decision is explained by response to competition as well as the temporal stability and persistence of response to competition.

Several advantages of our setting are worth noting. First, both the HSEE and the mock exams are administered in tightly controlled settings and grading is anonymous and carried out in a centralized location, giving us a lab-like level of control in the administration of the tasks. Second, in this context, feedback on relative performance is abundant and students have ample experience with the exam taking process. Students in China are exposed to exams throughout their educational career, and performance and rankings are made known after each exam. Mock exam results, for example, are returned within 5-7 days of taking the exam and contain scores and rankings. Therefore, the role of subjective beliefs about one's ability, or overconfidence, is arguably less likely to play a role in determining response to competition in this setting. Third, the students we study are under compulsory education, and therefore are not selected on ability. Because we have the universe of observations for each cohort, we observe close to the population of individuals within each birth cohort. Furthermore, specialization into tracks does not occur until high school, which means that response to competition and the other determinants of STEM choice in our analysis are observed prior to specialization, eliminating potential endogeneity to curriculum choice. Lastly, the characteristics of the county we study are highly similar to that of the median Chinese county, making our data arguably broadly representative of the Chinese population in this age group.

We find a 1 percentile point (2.7% standard deviation) gender difference in response to competition (i.e., scoring higher in the actual HSEE relative to the mock exam) favoring boys in STEM subjects. In contrast, in non-STEM subjects, the pattern is reversed, with girls exhibiting a 0.35 percentile point (1.3% standard deviation) higher response to competition than boys. All results are precisely estimated. This pattern of domain-specific gender differences can be observed across the ability spectrum. We rule out alternative explanations that the results are explained by teachers grading boys and girls differently, by parents caring more about sons' success, by boys putting in more effort when it counts, by gender differences in subjective beliefs, and by gender differences in response to feedback. In particular, we find no significant gender differences in performance improvement between two mock exams, in either domain.

Using principal component analysis on individual response to competition across all seven academic subjects, we find that the first principal component can be described as domain-general response to competition while the second principal component roughly corresponds to the comparative advantage in response to competition in STEM subjects over non-STEM subjects. Gender differences favor boys in both components.

Both response to competition in STEM and in non-STEM subjects have incremental significant explanatory power for choosing the STEM track over the non-STEM track in high school. The raw gender gap in STEM choice is 24.2 percentage points, which reduces to 18.3 percentage points when controlling for known predictors of STEM choice: STEM ability, comparative advantage in STEM, peer or role model effects, age, and SES. Response to competition in STEM and non-STEM together explains an additional 4 percentage points, or 22% of the remaining gender gap. Additionally, using the principal components as measures of response to competition, we find that both the domain-general component and the comparative advantage component are significant predictors of STEM choice. However, only the comparative advantage component contributes to the gender gap in STEM choice, explaining 19% of the adjusted gender gap, which further highlights the importance of domain-specific gender differences in response to competition.

Lastly, using the CEE and mock CEE scores taken three years after the HSEE, and defining response to competition in the CEE similarly as was done for the HSEE, we find that response to competition in the HSEE is significantly correlated with response to competition in the CEE. The size of the correlation coefficients are comparable to that found for risk preferences measured across a similar time horizon (Chuang and Schechter, 2015). We further find that response to competition in the CEE continues to favor male students in the STEM domain and female students in the non-STEM domain, with the gender gap intensifying over time. These represent some of the first evidence on the stability and persistence of response to competition over time.

Put together, these results suggest that an important contributor to the gender pay gap – field of study - is determined by domain-specific response to competition. Choosing to study STEM is not only explained by STEM ability, but also by one's ability to perform under competition, in particular in STEM over non-STEM subjects. If response to competition do not map consistently to labor market productivity - for example, if the level of competition encountered "on the job" bears little resemblance to the intensity of competition experienced during the HSEE - these results indicate a potential for job and occupational mismatch. More broadly, the results call for more careful consideration of the role that competition plays in selection mechanisms with significant welfare implications, including but not limited to high-stakes exams.

2 Related Literature

Our study follows others using educational competitions to study gender differences in performance in less versus more competitive exams. The consistent conclusion from these studies is that men outperform women in competitive exams relative to less competitive exams in settings as diverse as the French HEC entrance exam (Ors et al., 2013), the Graduate Record Exam in the United States (Schlosser et al., 2019), and in a regional extra-curricular math contest in Spain (Iriberri and Rey-Biel, 2019). Azmat et al. (2016) find that male students in a Spanish high school outperform female students in high stakes tests relative to low stakes tests.[4] In the study most similar to ours, Cai et al. (2019) find that men outperform women in the Chinese College Entrance Exam relative to the mock College Entrance Exam. With the exception of Iriberri and Rey-Biel (2019) which only looks at math performance, these studies do not find, as we do, substantially different patterns of gender gaps across STEM and non-STEM subjects. This discrepancy can potentially be accounted for by an omitted variable bias, induced by regression to the mean coupled with gender differences in the baseline (i.e., less competitive) exam scores. Our data show that neglecting to control for the baseline scores would result in a misleading interpretation of the direction of the gender gap in response to competition in non-STEM subjects.

[4] A related study by Morin (2015) finds that male performance in university improved relative to female performance as a result of increased competition resulting from a double cohort of students in Canada.

Several lab experimental studies have found that the gender gap in response to competition depends on the nature of the task. Using tasks based on existing gender stereotypes around task proficiency, Günther et al. (2010), Shurchkov (2012) and Cotton et al. (2013) find that while men have higher response to competition than women in quantitative and analytical tasks, in verbal tasks there is no gender difference in response to competition, or the gender gap reverses in favor of women. Our results demonstrate that the same patterns across STEM and non-STEM domains can be observed in field data from real world competitions. We additionally identify a domain-general component of response to competition, indicating that despite the strong domain-specificity of gender patterns in response to competition, individual response to competition is not purely a domain-specific phenomenon.

In linking response to competition measured in the HSEE to track choice in high school, our study relates to the literature on the external validity or external relevance of willingness to compete with regard to education and labor market outcomes: Buser et al. (2014) and Buser et al. (2017) show that lab measures of willingness to compete predict choosing prestigious and math-heavy tracks for Dutch and Swiss high school students. Willingness to compete also predicts higher salary expectations (Reuben et al., 2017), and higher starting salaries of MBA graduates (Reuben et al., 2019). Using a nationally representative sample in the Netherlands, Buser et al. (2024) find that both incentivized and unincentivized measures of willingness to compete predict completion of more education, choosing higher paying fields of study, and conditional on education, higher incomes. Our study contributes to this literature along several dimensions. First, we link field rather than lab measures of response to competition to subsequent educational choices. Second, we link performance response to competition to subsequent educational choices, whereas previous studies exclusively focused on willingness to compete. Lastly, we find that domain-general response to competition significantly predicts STEM choice, which supports the dominant view in this literature that STEM is a competitive field that attract those that respond well to competition across domains. However, that support is qualified by the fact that only the comparative advantage component of response to competition contributes to the gender gap in STEM choice, the domain-general component does not.

Our findings contribute to a large literature on the determinants of the gender gap in STEM choice (see Kahn and Ginther (2017) for a review). The underrepresentation of women in STEM fields is a policy concern for two main reasons: expanding the STEM workforce is thought to increase productivity and economic growth (Carnevale et al., 2011), and increasing women's participation in these highly lucrative fields would reduce the gender pay gap (Beede et al., 2011). An obvious factor in choosing to study STEM is ability in STEM. Interestingly, average gender differences in STEM ability are small (Hyde et al., 2008), although gender differences can be large at the right tail (Ellison and Swanson, 2010). More recently researchers have investigated the role of comparative advantage in STEM, since high ability girls and women who do well in STEM may still choose to study non-STEM subjects because they do even better in non-STEM. Valla and Ceci (2014), Breda and Napp (2019), Loyalka et al. (2017), Card and Payne (2021), and Goulas et al. (2022) find that comparative advantage in STEM has additional explanatory power for STEM choice, conditional on STEM ability. Peer effects and role model effects form a third category of investigation, motivated by the idea that the under-representation of girls and women in STEM could discourage them from participation in and of itself. There has been no consensus with respect to the sign of this impact, with Brenøe and Zölitz (2020) and finding that the proportion of peers who are female reduces female STEM participation, while Mouganie and Wang (2020) and Bostwick and Weinberg (2022) show the opposite. Carrell et al. (2010) find that having female instructors increases the likelihood of female students graduating with STEM degrees, while the meta-analysis conducted by de Gendre et al.

(2023) find a small positive average effect of same gender teachers, with a relatively wide distribution covering negative effects as well. Our study finds that in addition to these factors, response to competition matters for STEM choice and contributes to a substantial proportion of the gender gap in STEM choice.

Lastly, our study provides some of the first evidence on the stability and persistence of response to competition over time. While the stability of psychological traits has been widely studied, evidence on the stability of economic preferences is relatively sparse (Borghans et al., 2008). This literature is largely concentrated on the study of risk preferences, with less evidence on time preferences and social preferences (Chuang and Schechter, 2015; Schildberg-Hörisch, 2018). We find temporal stability in response to competition of a magnitude comparable to that found with risk preferences over a similar time horizon. Furthermore, response to competition measured three years later showed persistent and intensified domain-specific patterns of gender gaps. This illuminates the potential mechanisms by which response to competition measured pre-market contributes to long-run gender differences in labor market outcomes (Bertrand, 2011; Niederle and Vesterlund, 2011; Shurchkov and Eckel, 2018).

3 Institutional Background

In this section we provide an overview of the context and incentives involved in the academic exams taken at the end of middle school, and the choice of STEM or non-STEM track, made toward the beginning of high school.

3.1 High School Entrance Examination

Our measure of performance under competition uses scores from China's High School Entrance examination (HSEE), also known as *Zhongkao*. Administered by the education bureau in each province, the HSEE is a highly consequential and competitive exam which can only be taken once a year. While education up to the end of junior high school (Year 9) is compulsory, priority for admissions into high schools is essentially solely determined by the score on the HSEE.[5] Those who do not gain admissions into a high school are not eligible to apply for university. The Ministry of Education sets a 50% admissions rate for high schools as a guide,[6] which is followed closely in Henan Province, with an admissions rate of 51.8% in 2019.[7] High schools are further divided into elite and regular high schools, and the admissions rate for elite high schools in Henan Province is around 20% (see Appendix Table B.2 for the admissions rates of regular and elite high schools in the county where we collect data).

In some provinces, including Henan Province, the HSEE also serves as the junior high school graduation examination, which is used as a learning assessment of the nine years of compulsory education and is required for obtaining a diploma. The proportion of students taking the HSEE is consistently high throughout our sample years, around 90% from 2015-2016 and approaching 95% from 2017-2019 (see Appendix Table B.2).

3.2 Mock Examinations

Mock exams are held as practice for the HSEE. The format of the mock exam follows exactly that of the HSEE, in the subjects tested, the order of the subjects tested, and the amount of time allotted for each subject. See Appendix Table C.1 for the schedule of the HSEE by subject.

[5] Vocational high school, which does not allow for advancement into university, is another option for students. Although scores on the HSEE do not matter for admissions into vocational high school, student must have a record of having taken the exam to be admitted. We use the term "high school" to refer to academic high school exclusively.

[6] See the "National medium and long-term educational reform and development program (2010-2020)" accessible at http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/201008/t20100802_93704.html

[7] In 2019, there were 1.41 million students taking the HSEE in Henan Province, with a total high school capacity of 0.73 million. See http://m.xinhuanet.com/ha/2019-05/17/c_1124505191.htm

In Henan Province, two mock exams are given, with about one month's time between each mock exam and one month's time from the second mock exam to the HSEE (see Figure 1 for a timeline of the events).[8] The results of the mock exams (and the real exam) are released between 5 and 7 days after each exam. Within-county rankings are also released at this time.

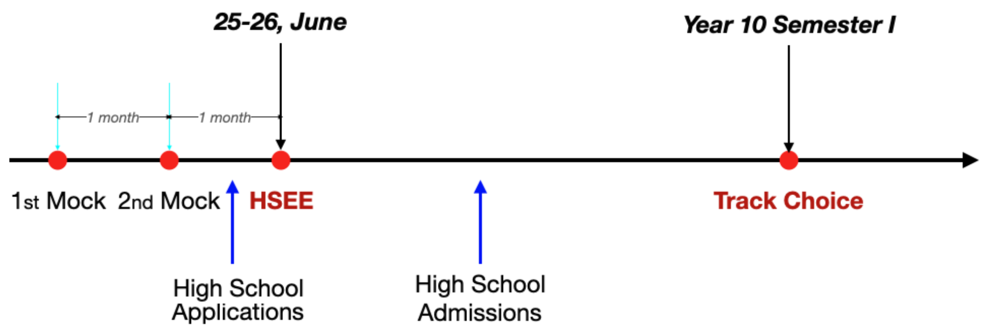


Figure 1: Timeline

While scores in the mock exams have no further consequences, students have an incentive to perform to the best of their ability because of the informational value of these exams for high school applications: After the second mock exam and prior to the HSEE, students submit their preference list for high schools, indicating, for example, whether their first choice is a regular high school or an elite high school. Admissions is based on the Boston Mechanism, with all high schools using the (total) HSEE score to determine priority (see Appendix C for details on the admissions procedure). The Boston Mechanism is well-known for not being strategy-proof (see, for example, Ergin and Sönmez (2006) and Abdulkadiroğlu et al. (2011)), so that both under- and over-estimating one's likelihood of acceptance into elite high schools are potentially costly. Empirically, it appears students take the mock exams seriously: Pearson correlation coefficients with the HSEE are around 0.95 for both the first mock exam and the second mock exam (see Table 2).

[8] See https://www.sohu.com/a/308138492_697551 for the 2019 mock exam schedule for the county where we collect data.

3.3 Choice of STEM vs. non-STEM Tracks

The requirement to choose a track in high school is common in many Asian and European countries.[9] In the county where we collect data, the choice of STEM or non-STEM track is made at the end of the first semester in high school, in Year 10 (refer to Figure 1). Prior to that, students study a common curriculum in high school. Once students choose their track, their curricula begin to differ. Track choice cannot be changed once chosen. The subjects common to both tracks are Chinese language, English Language, and Math, although Math content is easier in the non-STEM track. STEM track students will further study Physics, Chemistry, and Biology, while non-STEM track students will further study Politics, History and Geography. The subjects taught are also the subjects tested for in the College Entrance Exam (CEE).[10]

STEM jobs are more lucrative than non-STEM jobs, in China as it is elsewhere in the world Hu and Vargas (2015); Beede et al. (2011). Of more immediate concern to the students, the STEM track also affords more major options when applying for university.[11] In the university admissions process, students are admitted not just to a university but also to a specific major within the university. STEM track students are eligible to apply to 12 out of the 13 main university major categories in China, while non-STEM track students are eligible for about half of the 13 majors.[12] Universities in turn have major-specific quotas

[9]Countries like Germany and Singapore implement early student tracking into different educational paths based on ability, beginning at the end of elementary school (around ages 10-12) (Dustmann et al., 2017). In the Netherlands, students in secondary school pre-university programs must select one of four study tracks – science, health, social sciences, or humanities – at the end of their third year (age 15) in a six-year curriculum (Buser et al., 2014). In the United States, while there is no formal tracking in high schools, students aiming for university typically enroll in essential courses required for university admission. Research indicates a strong correlation between these preparatory courses in high school and subsequent university program choices (Card and Payne, 2021).

[10]In the CEE, Chinese Language, English Language, and Math are each worth 150 points. The subjects specific to each track are worth a combined 300 points, with students in each track being tested only on the subjects taught in their track.

[11]The same is true in Canada, for example (Card and Payne, 2021).

[12]STEM track students are eligible to apply to all majors except the Art major (category 13), while non-STEM track students can only apply to the Management (category 12), History (category 06), Literature (category 05), Education (category 04), Law (category 03) and some of the Economics (category 02) and

for each province, with more slots reserved for STEM track graduates than non-STEM track graduates.[13] Correspondingly, substantially more students choose the STEM track. In the county where we collect data, around two-thirds of students choose the STEM track, and the university admissions rate is similar across the two tracks, at between 80-85%.[14]

4 Data

The county where the data are collected is in Henan province, in central China. The characteristics of the county is similar to that of the ”median” Chinese county. Along the dimensions of GDP per capita, average years of education, population density, and urbanization rate, the values for our sample county closely align with the median values across all Chinese counties (as shown in Appendix Figure A1). With a population of 350,000 in 2010, our county is also close in size to the median Chinese county, which has a population of 380,000, according to the 2010 Population Census.

From administrative sources we collected the universe of observations in each year for the data categories listed below:

- **HSEE scores by subject (2015-2019)**, along with gender, birth date, middle school name, class identifier, student name and national ID.
- **HSEE 2nd mock exam scores by subject (2015-2016, 2018-2019)**, along with middle school name, class identifier, and student name.

Philosophy (category 01) majors.

[13] For high school graduates in Henan province for instance, in 2019 there were 374 Tier 1 universities offering 78,045 slots for STEM track graduates and 16,157 slots for non-STEM track graduates. Similarly, there were 851 Tier 2 universities offering 134,328 slots for STEM track graduates and 61,274 slots for non-STEM graduates. These figures were obtained from the Higher Education Examination Authorities of Henan Province website accessible at http://www.haeea.cn/adc/pzlqtj.shtml.

[14] Using data from Ningxia province, Loyalka et al. (2017) find that choosing the STEM track increases university enrollment by 19 percentage points, in an instrumental variables regression where STEM choice is instrumented by comparative advantage in STEM, conditional on total HSEE score.

- **HSEE 1st mock exam scores by subject (2018-2019)**, along with middle school name, class identifier, and student name.

- **CEE scores by subject (2018-2021)**, along with gender, birth date, high school name, class identifier, student name, national ID, and track.

- **CEE mock exam scores by subject (2018-2019)**, along with high school name, class identifier, and student name.

- **Family information from Poor Households Registration and Management System (PHRMS) (2015-2019)**

  The data source consists of the universe of observations on households and their members qualifying as ”poor” in the sample county. In the sample county there are around 12,000 poor households with 45,000 poor members. Each individual in this dataset can be uniquely identify by their national ID.

We next undertook a large-scale data merging process, producing the following merged datasets used in the empirical analysis:

1. **HSEE scores and HSEE mock exam scores**, along with gender, birth date, poor/non-poor status (from PHRMS), middle school class identifier, and middle school characteristics for the universe of HSEE takers from 2015-2016 and 2018-2019.

2. **HSEE scores, HSEE mock exam scores, and track choice**, along with gender, birth date, poor/non-poor status (from PHRMS), middle school class identifier, middle school characteristics, high school class identifier, and high school type (elite/regular) for the universe of high school students starting high school in 2015, 2016, and 2018.

3. **HSEE scores and HSEE mock exam scores, CEE scores and CEE mock exam scores**, along with gender, birth date, poor/non-poor status (from PHRMS), track, and high school type (elite/regular) for the universe of CEE takers in 2018 and 2019.

Note that the HSEE, CEE, and PHRMS data are merged to each other using the individual's national ID, which is a unique identifier. Mock exams (both HSEE and CEE) and track choice data do not contain national IDs, but contain each exam-taker's name, school, and class identifier, which we use to merge to the other data sources.[15] For further details on the merging process see Appendix B.

In and out-migration is of minimal concern in this context. China's unique Hukou system ensures that the vast majority of students pursue their education within their Hukou-registered county.[16] Data from the national student enrollment management system shows that over 98% of elementary school graduates in the county who enrolled in middle schools remained in the county, as did over 99% of middle school graduates in the county who continued on to high school. This ensures that our dataset encompasses virtually all students who are Hukou-registered in the county and minimizes data attrition during the transition from middle to high school.[17]

Descriptive statistics on each of the three merged datasets are reported in Appendix Appendix Table B.1, which uses only matched observations, and reproduced in Appendix Table B.1 using all available years of data.

Table 2 reports descriptive statistics for our main analysis by gender, for the cohorts taking

[15]98.36% of the mock HSEE data were successfully matched to HSEE data. 97.99% of the mock CEE data were successfully matched to CEE data.

[16]Per the Ministry of Education's guidelines, children of compulsory education age are required to enroll in schools close to their Hukou-registered addresses. This rule is particularly stringent for academic high school students, as their registration and admission into college entrance examinations are closely tied to their Hukou status, necessitating that students register for and take these exams in their Hukou-registered area. This system often results in children remaining in their Hukou-registered areas while their parents migrate to larger cities for work, leading to the phenomenon of 'left-behind children' (Li et al., 2015).

[17]The national student enrollment management system serves as a centralized repository tracking the educational enrollment status of students across the country, from elementary through to academic high school. The system provides detailed records of the schools and classes from which students have graduated and the institutions they subsequently enrolled in. Using data derived from this system for our sample county, we are able to observe all students who graduated from institutions within the county at each educational stage.

Table 1: Descriptive Statistics

Variable	Mean	S.D.
Overall sample		
1. Middle school		
Years available: 2015, 2016, 2018, 2019		
Observations: 19,549		
Total score in HSEE	326.5	111.0
STEM subjects score in HSEE	118.5	58.1
Non-STEM subjects score in HSEE	208.0	56.8
Total score in 2nd mock	320.1	102.7
STEM subjects score in 2nd mock	116.4	52.0
Non-STEM subjects score in 2nd mock	203.7	54.8
Female dummy	0.46	0.50
Age at HSEE	15.60	0.70
Urban middle school dummy	0.61	0.49
Poverty status dummy	0.13	0.34
2. High school		
Years available: 2015, 2016, 2018		
Observations: 6,458		
STEM track dummy	0.66	0.47
Elite high school dummy	0.48	0.50
Female dummy	0.50	0.50
Age at HSEE	15.48	0.68
Urban middle school dummy	0.75	0.43
Poverty status dummy	0.10	0.30
3. College Entrance Exam		
Years available: 2018,2019		
Observations: 3,764		
Total score in CEE	441.9	87.6
STEM subjects score in CEE	197.9	97.2
Non-STEM subjects score in CEE	244.0	81.4
Total score in 4th mock	418.9	90.37
STEM subjects score in 4th mock	176.4	97.6
Non-STEM subjects score in 4th mock	242.5	79.6
Elite high school dummy	0.53	0.50
Female dummy	0.53	0.50
Age at CEE	18.2	0.68
Poverty status dummy	0.10	0.31

the HSEE from 2015-2016 and 2018-2019. The sex ratio of male to female students in the sample is 1.15, which is close to that of the sex ratio in the population.[18] Average age is between 15 and 16 for both boys and girls. Around 60% are from an urban middle school, with boys somewhat more likely to be from an urban middle school, and about 13% come from a poor family, with girls slightly more likely to be poor.

Panel B of Table 2 reports the means of the raw scores on the 2nd mock exam and the HSEE by gender, aggregated across the four cohorts. On the 2nd mock exam, while girls and boys perform equally well in STEM subjects, girls perform significantly better than boys in non-STEM subjects. Taking the point estimates at face value, both boys and girls improved their performance in the HSEE, with boys improving more than girls in STEM, and girls improving slighty more than boys in non-STEM. However, the comparison of the raw scores over the four cohorts cannot be taken at face value because the exams in each cohort are written by different sets of educators and can vary in content and difficulty, such that the same student taking exams from different years would likely have different scores. For the main empirical analysis we recode the raw scores into percentiles and standardized scores.

Conditional on having taken the 2nd mock exam, the rate of taking the HSEE is nearly universal for both boys and girls. The Pearson correlation coefficients between performance in the mock exams and performance in the HSEE are around 95% for both genders, for both mock exams. For the rest of the paper, unless otherwise specified, the term ”mock exam” without qualifiers will refer to the second mock exam.

[18]According to the National Bureau of Statistics (2012), the sex ratio at birth stood at 113.8 in 1990 and increased to 119.9 by 2000. The cohort taking the HSEE in 2015 would have been born around the year 2000.

Table 2: Descriptive Statistics by Gender

	Male			Female		
Explanation	Obs	Mean	S.D.	Obs	Mean	S.D.
Panel A. Demographics						
Age at HSEE	10,453	15.62	0.68	9,096	15.57	0.72
Urban middle school dummy	10,453	0.64	0.48	9,096	0.59	0.49
Poverty status dummy	10,453	0.12	0.33	9,096	0.14	0.35
Panel B. Exam performance						
STEM subjects score in 2nd mock	10,453	116.4	54.9	9,096	116.4	48.5
Non-STEM subjects score in 2nd mock	10,453	193.6	57.2	9,096	215.3	49.5
STEM subjects score in HSEE	10,453	119.3	60.5	9,096	117.6	55.1
Non-STEM subjects score in HSEE	10,453	197.8	57.9	9,096	219.7	53.1
% taking HSEE, cond'l on taking mock exam	10,453	0.988		9,096	0.996	
Correlation btwn 2nd mock and HSEE	10,453	0.957		9,096	0.949	
Correlation btwn 2nd mock and HSEE (rank)	10,453	0.961		9,096	0.950	
Correlation btwn 1st and 2nd mocks	5,010	0.963		4,256	0.955	
Correlation btwn 1st and 2nd mocks (rank)	5,010	0.967		4,256	0.957	

5 Response to competition

5.1 Empirical Strategy

In order to meaningfully aggregate exam performance across cohorts, we recode raw scores of student i in cohort c in subject s into relative scores using two conventional methods from the education literature. The first method uses percentile ranks: R^s_{ic} equals 100 times the percentage of raw scores that are equal to or below that of student i in subject s in cohort c. The second method uses standardized scores: $P^s_{ic} = \frac{r^s_{ic} - \overline{r^s_c}}{\sigma_{r^s_c}}$ where r^s_{ic} is the raw score of student i in subject s, $\overline{r^s_c}$ is the average raw score in subject s in cohort c, and $\sigma_{r^s_c}$ is the standard deviation of raw scores in subject s in cohort c.

Our preference is for percentile rank coding, because priority in the admissions process is determined by ranks and therefore it is the more relevant margin on which students are competing. For robustness we report results using both coding methods.

Similar to Cai et al. (2019), the main estimation equation is

$$RC^s_{ic} = \alpha + \beta Female_{ic} + \gamma Mock^s_{ic} + \Theta \mathbf{X}_{ic} + \delta_c + \varepsilon^s_{ic} \tag{1}$$

where RC^s_{ic} is response to competition of student i in cohort c in subject s. When using percentile rank coding, $RC^s_{ic} = R^{Z,s}_{ic} - R^{M,s}_{ic}$, where the superscript Z indicates the HSEE and M indicates the mock exam. When using standardized score coding, $RC^s_{ic} = P^{Z,s}_{ic} - P^{M,s}_{ic}$. the variable of interest is $Female_{ic}$. $\mathbf{X_{ic}}$ a vector of controls consisting of age and SES, as proxied by whether the student comes from an urban school and whether the student comes from a poor family according to records in the PHRMS. δ_c are cohort fixed effects.

Leaving out $Mock^s_{ic}$ as a control for the moment, note that Equation (1), as pointed out by Cai et al. (2019), is the first differenced equivalent of a panel model with individual fixed effects, where performance in the HSEE and the mock exam are stacked and β is the coefficient on the interaction between *Female* and performance in the HSEE.[19] The implied inclusion of individual fixed effects indicate that individual heterogeneity that do not change between the mock and the real HSEE has been accounted for in Equation (1).

It is crucial to additionally control for performance in the mock exam, $Mock^s_{ic}$. Due to regression to the mean, performance in the mock exam will have a mechanical, negative relationship with improvement from the mock exam to the HSEE.[20] Not controlling for $Mock^s_{ic}$ will lead to an omitted variables bias on β whenever $Mock^s_{ic}$ differs by gender. In particular, because girls score higher than boys in non-STEM on the mock exam (see Table 2), $Mock^{non-STEM}_{ic}$ will be positively correlated with $Female_{ic}$ and negatively correlated with $RC^{non-STEM}_{ic}$, resulting in a negative omitted variable bias on β when $Mock^s_{ic}$ is left out. For this reason, in the subsequent results discussion, we only interpret findings on

[19]The panel set up is used in Azmat et al. (2016) and Iriberri and Rey-Biel (2019).

[20]A negative relationship between performance in the mock exam and performance improvement from the mock exam to the HSEE can also arise from the censoring of scores at 0 and the maximum score.

response to competition after having controlled for mock exam performance.

5.2 Baseline results

Figure 2 graphs the distribution of residualized response to competition, after controlling for $Mock^{s}_{ic}$ and cohort fixed effects. It clearly shows that boys have higher response to competition in STEM while girls have higher response to competition in non-STEM subjects. In the rest of the paper, ”residualized” response to competition will refer to response to competition after controlling for mock exam performance in the relevant domain and cohort fixed effects.

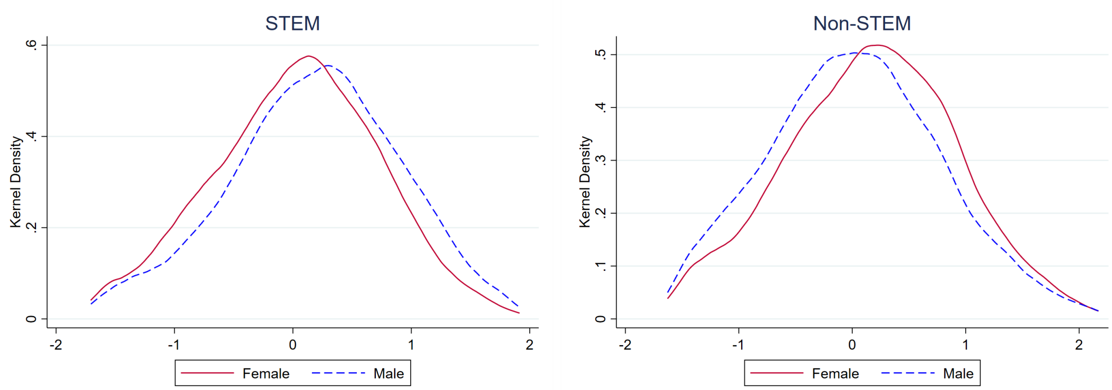


Figure 2: Distribution of residualized response to competition, using percentile ranks

Table 3 reports the estimation results of Equation 1 for STEM and non-STEM subjects separately. Columns (1) and (4) omit the control for *Mock*, i.e., performance in the mock exam, in order to demonstrate the impact of not correcting for the mechanical relationship between *Mock* and response to competition. In columns (2) and (5) we include the control for *Mock*, and the coefficients on *Female* can be interpreted as gender gaps in response to competition. Columns (3) and (6) additionally control for age when the HSEE was taken and SES. Panel A codes response to competition and mock exam performance using percentile ranks while Panel B uses standardized scores.

Consistent with the graphical evidence in Figure 2, Columns (2)-(3) and (5)-(6) in both Panels indicate that boys have higher response to competition in STEM subjects, while girls have higher response to competition in non-STEM subjects, with the magnitude of the gender effect sizes about twice as large for STEM subjects. The results in Panel A show a male advantage in STEM response to competition of 1 percentile point and a female advantage in non-STEM competitive outperformance of 0.35 percentile points. As a point of reference, using performance in the quantitative section of the actual GRE relative to performance on experimental GRE questions, Schlosser et al. (2019) find a male advantage of 3.9 percentile points in the real exam.

In Panel B, the gender gap in response to competition is 2.7% of a standard deviation in STEM, favoring boys, and 1.3% of a standard deviation in non-STEM, favoring girls. These are small effect sizes but non-trivial, as they are comparable to the impact of educational interventions using increased instruction time and teacher quality enhancement.[21] They are somewhat smaller than the gender gaps in response to competition found in other educational settings.[22] In the setting closest to ours, Cai et al. (2019) find a gender gap in response to competition of 4.8% of a standard deviation favoring boys in the College Entrance Exam versus a mock College Entrance Exam in China.

The comparison of the *Female* coefficients in the preferred specifications in Columns (2)-(3) and (5)-(6) to that in columns (1) and (4) demonstrates the importance of controlling

[21] Bietenbeck and Collins (2023) find that a one-hour increase in weekly instruction time can enhance achievement by 1.4%-2% of a standard deviation in PISA scores (using data from the 2015 and 2018 PISA waves, which are the years that overlap with the cohorts in this study). Using 2015 TIMSS data, Sancassani (2022) show that improving teacher quality through subject-specific qualification in a given subject can increase student test scores by 3.5% of a standard deviation in that subject.

[22] Azmat et al. (2016) found a gender gap of 5.1% of a standard deviation favoring boys in high-stakes math exams versus low-stakes math exams taken earlier in the year in Spain. Iriberri and Rey-Biel (2019) find a gender gap of 9% of a standard deviation favoring boys in the second and more competitive stage of a math contest versus the first stage of the same contest in Spain. Morin (2015) finds a gender gap of 11% of a standard deviation favoring boys in academic performance in a more competitive cohort relative to a less competitive cohort in Canada.

for *Mock* to correct for the mechanical relationship between response to competition and mock exam performance. The *Female* coefficient in Column (4), for non-STEM subjects, is of the opposite sign and statistically significant, consistent with the prediction of a negative omitted variable bias arising from the fact that girls do better in the mock exam in non-STEM subjects. Had we not controlled for *Mock*, we would have concluded that girls also have lower response to competition in non-STEM. On the other hand, the coefficient on *Female* in Column (1), for STEM subjects, is largely similar to that in the preferred specifications in Columns (2)-(3). This is due to the fact that girls and boys do similarly well in the mock exam in STEM subjects so that the omitted variable bias is much less prominent. These comparisons also suggest a possible explanation for why previous studies looking at response to competition in the educational setting, which did not employ the correction by controlling for performance in the non- or less-competitive setting, found that the gender gap in response to competition similarly favor boys in non-STEM subjects (see Ors et al. (2013), Azmat et al. (2016), Cai et al. (2019), Schlosser et al. (2019)).

The similarity of coefficients between columns (2) and (3) and between columns (5) and (6) suggest that age and socioeconomic status are not major drivers of the gender gaps in response to competition in either STEM or non-STEM subjects.

5.3 Robustness

5.3.1 Additional Controls

The baseline results are robust to controlling for middle school class size to proxy for teacher attention (Appendix Table D.1) and to allowing for the impact of regression to the mean to differ by gender by adding an interaction term between *Mock* and *Female* to Equation (1). The latter results are reported in Appendix Table D.3, where we find the average marginal effect of *Female* to be very similar to that in the baseline results.

Table 3: Response to competition (HSEE)

VARIABLES	(1)	(2) STEM subjects	(3)	(4)	(5) Non-STEM subjects	(6)
Panel A: Percentile rank						
Dep Var = RC^{STEM}				**Dep Var =** $RC^{Non-STEM}$		
Female	-0.902***	-0.906***	-0.972***	-0.321**	0.387***	0.430***
	(0.152)	(0.150)	(0.150)	(0.139)	(0.143)	(0.143)
$Mock^{\text{STEM}}$		-1.919***	-2.008***			
		(0.060)	(0.071)			
$Mock^{\text{Non-STEM}}$					-1.760***	-1.919***
					(0.061)	(0.070)
Constant	0.757***	0.759***	14.145***	0.503***	0.164	10.180***
	(0.174)	(0.171)	(1.845)	(0.163)	(0.160)	(1.705)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,549	19,549	19,549	19,549	19,549	19,549
R-squared	0.002	0.035	0.038	0.001	0.032	0.034
Panel B: Standardized score						
Dep Var = RC^{STEM}				**Dep Var =** $RC^{Non-STEM}$		
Female	-0.022***	-0.023***	-0.025***	-0.014***	0.013**	0.016***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$Mock^{\text{STEM}}$		-0.069***	-0.072***			
		(0.002)	(0.003)			
$Mock^{\text{Non-STEM}}$					-0.067***	-0.076***
					(0.003)	(0.004)
Constant	0.013**	0.014**	0.465***	0.011*	-0.002	0.435***
	(0.006)	(0.006)	(0.063)	(0.006)	(0.006)	(0.062)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,549	19,549	19,549	19,549	19,549	19,549
R-squared	0.001	0.036	0.039	0.000	0.034	0.039

Note: OLS regressions. The dependent variable is response to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the second mock exam. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

5.3.2 Subsample analysis by ability

Because high school admissions is limited to roughly 50% of students, the above median students may take the HSEE more seriously, and thus results from this subsample may be more reliable. In Appendix Table D.4 we limit the sample to those who scored in the top 50th percentile in the second mock exam. The results show that the gender patterns of response to competition across STEM and non-STEM are even more pronounced.

We further generalize the exercise by dividing the sample into 20 bins of 5 percentile points each, using the second mock exam, and separately estimate Equation (1) for each bin. The results are displayed in Figure 3. With the exception of the lowest ability bin, there is a clear separation of the *Female* coefficients for STEM and non-STEM subjects, across the ability spectrum. The coefficients are consistently negative for STEM and positive for non-STEM although the 95% confidence intervals cross zero for some bins in the lower part of the ability spectrum in non-STEM and in the lowest bin in STEM.

These findings suggest that the domain-specific patterns of gender gaps in response to competition are a pervasive phenomenon not driven by certain segments of the ability spectrum. This adds to the growing evidence on the gender gap in response to competition across the ability spectrum, contrasting with much of the literature that focuses on high ability populations.[23]

5.4 Mechanisms

5.4.1 Differential grading by teachers

Teachers have been known to grade boys and girls differently (Lavy, 2008; Lavy and Sand, 2018; Lavy and Megalokonomou, 2019; Terrier, 2020). In order for this tendency to explain

[23] See Buser et al. (2024) and references therein for studies on willingness to compete using representative samples.

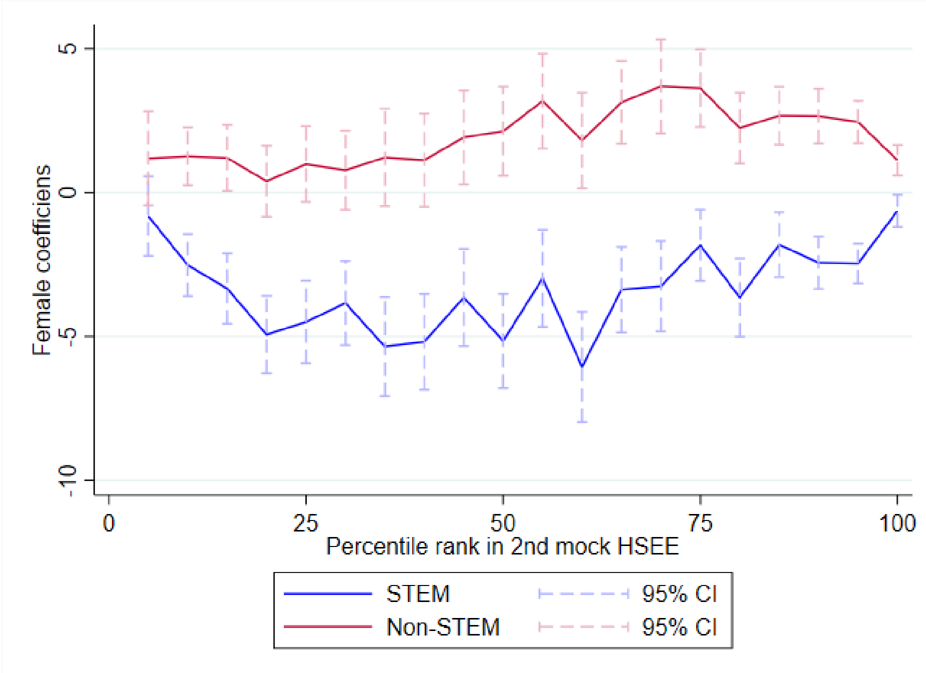


Figure 3: Gender gaps in response to competition by ability

the results, teachers would need to be able to identify the gender of the exam-taker, and additionally grade boys and girls differently in different subjects, on mock versus real exams. In our context, this is virtually impossible due to the anonymous and centralised grading process in this province, which is identical for the mock exam and for the HSEE. Grading is carried out centrally at a closed location by each prefectural educational bureau. The graders are selected from middle and high school teachers across the prefecture, with teachers of Year 9 students (the cohort taking the HSEE) excluded. Exam papers are scanned, after which the response to each constructed response question is cut out and collated with the responses of all the other test-takers to the same question. Any given constructed response question is graded by the same set of graders, who have no information on the student's identity or their responses to other questions in the exam. Multiple choice questions are machine-graded.[24]

[24]See http://lywb.lyd.com.cn/html2/2018-05/15/content_161611.htm and https://baijiahao.baidu.com/s?id=1637850151105649463&wfr=spider&for=pc for more details on the grading process.

5.4.2 Differential parental involvement

China is a country with a large sex ratio, indicative of son preference (Li, 2007; Peng, 2011). Parents, therefore, may care more about the success of sons rather than daughters in the HSEE, and may push their sons harder. In order for this factor to influence the results, parental involvement must be effective in incrementally improving performance in the one month period between the mock exam and the HSEE.[25] Putting aside the issue of whether parents are able to do so, especially given that the education of the average individual in this county is 8.4 years, below that of middle school completion (see Appendix Figure A1 Panel b), if greater parental involvement on the HSEE for boys is a driving factor, we would expect to find a male advantage in response to competition in all subjects, since it is the total HSEE score that determines priority in high school admissions. This is inconsistent with the male disadvantage in response to competition in non-STEM subjects. Furthermore, if SES can be taken as a measure of parental capacity to influence HSEE outcomes, Columns 3 and 6 in Tables 4 and 5 show that controlling for parental capacity has little impact on the magnitude of the gender patterns in response to competition across domains.

5.4.3 Boys do not take the mock exam seriously

Another potential explanation for the patterns we find is that boys do not take the mock exam seriously. For example, mens' higher performance than women in the real GRE compared with their scores on experimental GRE questions can be attributed to their lower effort in experimental GRE questions (Schlosser et al., 2019). However, unlike experimental GRE questions, which have no instrumental value to the test-taker, mock exam performance, as previously noted, has important informational value for the high school application process and is unlikely to be not taken seriously. Consistent with this

[25] As we will show in Section 5.4.5, there is no gender difference in performance improvement between the two mock exams, in either domain.

observation are the high correlations between each mock exam and the HSEE, for both boys and girls, reported in Table 2. Furthermore, given the domain-specific patterns of gender differences in response to competition, a simple story of one gender taking mock exams less seriously is not a sufficient explanation for the results.

Evidence from other studies further casts doubt on this alternative explanation in the Chinese context. Gneezy et al. (2019) gave students in China and in the US a PISA-like test. Treatment group students were given an incentive for good performance while the control group had no incentives. The Chinese students performed equally well whether in the control or treatment group (and as well as the best US students), while the US students' performance improved significantly with the incentive. The authors conclude that unlike students in the US, students in China perform to the best of their abilities even in low-stakes environments.[26] In the setting closest to ours, Cai et al. (2019) conduct a survey of high school students in China as part of their study on response to competition in the College Entrance Exam, and find no gender differences in preparation for the mock College Entrance Exam, or the real College Entrance Exam, in any subject.

5.4.4 Boys are more overconfident than girls

Psychology and behavioral economics research show that men are more overconfident than women (Barber and Odean (2001), Croson and Gneezy (2009)), especially in stereotypically male domains (Coffman, 2014; Bordalo et al., 2019). Could this explain our results, presuming that boys are more overconfident in STEM than non-STEM subjects? First, it's not clear that overconfidence should lead to better performance - as it can also result in the belief that less effort is required to do well (Santos-Pinto and Sekeris, 2023). Second, as mentioned earlier, in this setting, students receive exhaustive feedback regarding their absolute and relative academic performances, both in their regular classes and after each

[26]And that this potentially contributes to the China-US gap in performance in the PISA, which is an unincentivized test.

mock exam (and after the HSEE). In particular, within 5 to 7 days after each mock exam, students are told their scores and their school and county-wide ranks in each subject. It is therefore difficult for students to hold on to mistaken beliefs about their relative abilities in this setting.[27] Lastly, Appendix Table D.2 reports results of estimating Equation (1) with an added control for within-class rank on the mock exam as a proxy for confidence (see Murphy and Weinhardt (2020)).[28] The pattern remains that response to competition favors boys in STEM and favors girls in non-STEM and is, if anything, more pronounced.

5.4.5 Boys and girls respond differently to the same feedback

The previous section established that boys and girls receive equally informative and accurate feedback from the mock exams. However, they could respond differently to the same feedback - a large body of literature finds that girls are more sensitive to negative feedback than boys (Roberts and Nolen-Hoeksema, 1989; Rask and Tiefenthaler, 2008; Ost, 2010; Buser, 2016). Furthermore, Coffman et al. (2023) show that beliefs updating after receiving positive feedback is biased toward gender-congruent domains. To investigate this alternative interpretation, we exploit the fact that there are two mock exams, both non-competitive, also held one month apart. If there is differential response to feedback that occurs during a one-month period, we should also see a similar pattern of gender differences in performance differences between the two mock exams.

We conduct this placebo test by replacing the outcome variable in Equation (1) with the difference between the second and the first mock exam. The results are reported in Table 4. Contrary to our baseline results on response to competition in Table 3, we find no

[27]Zhang (2019) finds, as part of a Niederle-Vesterlund style experiment, that high school students in China were on average accurate in their incentivized guesses on their own rank in a four-person tournament based on a real-effort task, and there was no gender difference in the accuracy of guesses, which the author attributes to the unique availability of relative feedback in schools in China. In contrast, Niederle-Vesterlund style experiments elsewhere in the world typically finds men to be more confident than women (Niederle and Vesterlund, 2011).

[28]In our setting the within-class rank is arguably a measure of *over*confidence given that the regression also controls for overall rank in the mock exam.

significant gender differences in performance improvement between the two mock exams, in either STEM or non-STEM subjects, and the magnitudes of the gender coefficients are substantially smaller. Appendix Table D.5 finds the same results when limiting the data to the above median subsample.[29]

5.4.6 Stereotype threat

Given the insufficient support for the alternative explanations, the results suggest that the gender differences in performance improvement observed between the mock exam and the HSEE reflect gender differences in response to the competitive pressure of the HSEE, per se. This is consistent with stereotype threat, which is the phenomenon that negative stereotypes can reduce the performance of the stigmatized group (Steele and Aronson, 1995; Spencer et al., 1999). The stereotype, in both explicit and implicit form, that math and science are for males and humanities and liberal arts for females is widely held across societies around the world (Nosek et al., 2007; Miller et al., 2014)[30] and may originate in the gender composition at the right tail in each domain (Bordalo et al., 2016). Such stereotypes, when held by teachers, has been shown to impact student performance (Carlana, 2019). In framework developed in Dee (2014), the stereotype threat effect in our context can be conceptualized as a negative ability shock. A review of the literature finds that the impact of stereotype threat is stronger "when stress levels are more extreme" (Schmader et al., 2008), which suggests that the size of the negative ability shock will be larger in the HSEE compared to the mock exam.

[29] Data availability on the 1st mock exam is confined to the years 2018-2019, resulting in a sample size roughly half of that in the baseline findings. To ensure robustness, we present the baseline outcomes restricted to these two cohorts (Appendix Table D.6), as well as restricted to the above median subsample in these two cohorts (Appendix Table D.7). The results closely mirror the baseline findings, with the exception of the insignificant gender difference in non-STEM for the overall sample.

[30] In particular, these stereotypes (both implicit and explicit) have been documented among Chinese middle and high school students (Liu et al., 2010).

Table 4: Performance improvement between the mock exams

	STEM subjects			Non-STEM subjects		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Percentile ranks						
Female	-0.046	-0.062	-0.147	0.183	-0.215	-0.217
	(0.186)	(0.185)	(0.199)	(0.188)	(0.196)	(0.210)
$Mock^{\text{STEM}}$		-0.819***	-1.174***			
		(0.084)	(0.099)			
$Mock^{\text{Non-STEM}}$					-0.904***	-0.960***
					(0.095)	(0.108)
Constant	-6.045***	-5.871***	-12.408***	-5.747***	-5.396***	-8.601***
	(0.166)	(0.164)	(2.220)	(0.172)	(0.173)	(2.318)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,266	9,266	9,266	9,266	9,266	9,266
R-squared	0.096	0.103	0.111	0.080	0.089	0.090
Panel B: Standardized Score						
Female	0.005	0.005	0.002	0.011*	-0.009	-0.009
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)
$Mock^{\text{STEM}}$		-0.023***	-0.036***			
		(0.003)	(0.004)			
$Mock^{\text{Non-STEM}}$					-0.047***	-0.050***
					(0.005)	(0.006)
Constant	-0.214***	-0.209***	-0.471***	-0.196***	-0.178***	-0.340***
	(0.006)	(0.006)	(0.077)	(0.006)	(0.006)	(0.083)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,266	9,266	9,266	9,266	9,266	9,266
R-squared	0.097	0.102	0.110	0.068	0.085	0.086

Note: OLS regressions. The dependent variable is the 2nd mock exam score minus the 1st mock exam score. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the first mock exam. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

5.5 Domain-general and comparative advantage in response to competition

The gender gaps in response to competition observed in the previous section vary significantly across STEM and non-STEM subjects. However, this does not negate the existence of domain-general response to competition. We explore the decomposition of response to competition using principal component analysis in this section

Table 5 reports the correlation between (residualized) response to competition in STEM and (residualized) response to competition in non-STEM. We find positive and significant correlation coefficients for the full sample, as well as separately for boys and girls. Controlling for age and SES has virtually no impact on the strength of the correlation.[31] This indicates that a student who responds positively to competition in STEM tends to also respond positively to competition in non-STEM. Appendix Table D.8 reports the pairwise correlation coefficients for all seven academic subjects separately. Response to competition in all subjects is positively and significantly correlated with response to competition in every other subject.

We next carry out a principal component analysis (PCA) of the (residualized) response to competition in each of the seven subjects. The first principal component accounts for 33% of the variation in individual response to competition, with all seven subjects loading positively onto this component (see Table 6). We interpret this as the domain-general component of response to competition. This is similar to the approach taken by Dohmen et al. (2011), who interpret the first principal component of attitudes toward risk across different domains as the domain-general component of risk preferences. On the second principal component, which explains 14% of the variation, the factor loadings are positive for the

[31] Precisely, to obtain correlation coefficients, we regress the standardized (residualized) response to competition in STEM on the standardized (residualized) response to competition in non-STEM. To obtain correlations with controls, we first residualize standardized response to competition on the controls.

Table 5: Correlation in Response to Competition between STEM and non-STEM

	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample		Males		Females	
Panel A: Percentile ranks						
	0.342***	0.344***	0.306***	0.308***	0.386***	0.387***
	(0.010)	(0.010)	(0.012)	(0.012)	(0.015)	(0.015)
Panel B: Standardized scores						
	0.348***	0.351***	0.301***	0.305***	0.414***	0.416***
	(0.010)	(0.010)	(0.012)	(0.012)	(0.015)	(0.015)
Controls	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,549	19,549	10,453	10,453	9,096	9,096

Notes: Correlation coefficients between residualized response to competition in STEM and non-STEM. Columns 1-2 report results for the full sample, Columns 3-4 report results for the male subsample, and Columns 5-6 report results for the female subsample. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

non-STEM subjects and negative for the STEM subjects, with the exception of English, which has a small but negative factor loading. To ease exposition, we flip the sign on the second principal component and interpret it as the comparative advantage component of response to competition: it captures the extent to which there is a comparative advantage in response to competition in STEM over non-STEM subjects. These patterns of factor loadings additionally validate the categorization of subjects into STEM and non-STEM in our baseline analysis.

Table 7 reports the gender difference in these two components. We find that boys exhibit significantly higher domain-general response to competition, and significantly higher comparative advantage in response to competition. The latter finding is unsurprising given boys' advantage (disadvantage) in response to competition in STEM (non-STEM)

Table 6: Factor Loadings in PCA

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	STEM subjects			Non-STEM subjects			
	Maths	Physics	Chemistry	Chinese	English	History	Politics
PC_1	0.52	0.42	0.43	0.33	0.36	0.36	0.32
PC_2	-0.30	-0.40	-0.30	0.51	-0.07	0.29	0.55

Notes: This table presents the factor loadings on the first and second principal components in a PCA of residualized response to competition using all seven subjects. Scores are coded using percentile ranks.

subjects. Gender differences are more pronounced for the comparative advantage component than the domain-general component when considering only the above-median subsample, which is the more relevant group for analyzing determinants of STEM choice in high school. Appendix Table D.9 reproduces the analysis for urban and rural middle school students separately, and Appendix D.10 for students from non-poor and poor families. The findings are consistent with those of the full sample, with gender differences favoring boys in both the domain-general component and in the comparative advantage component. Though for students from rural middle schools and from poor families, the gender difference in the domain-general component is insignificant.

This is the first attempt, to our knowledge, to decompose response to competition into domain-general and comparative advantage components. We find support for the existence of both components. The results also suggest that important education and labor market outcomes could depend differently on domain-general versus comparative advantage in response to competition, which we investigate in the next section.

6 STEM track choice

In this section we test whether response to competition explains the decision to pursue the STEM track over the non-STEM track in high school. Track choice, as previously described, is a once-and-for-all choice made in Year 10, the first year of High School. Figure 4 shows

Table 7: Gender gaps in domain-general and comparative advantage in response to competition

	Overall		Above median		Below median	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: PC_1						
Female	-0.072***	-0.081***	-0.049*	-0.069**	-0.099***	-0.097***
	(0.021)	(0.021)	(0.029)	(0.029)	(0.030)	(0.030)
R-squared	0.001	0.005	0.001	0.007	0.002	0.007
Panel B: $-PC_2$						
Female	-0.093***	-0.097***	-0.140***	-0.148***	-0.045**	-0.053**
	(0.014)	(0.014)	(0.020)	(0.020)	(0.021)	(0.021)
R-squared	0.002	0.006	0.005	0.007	0.001	0.010
Observations	19,549	19,549	9,883	9,883	9,666	9,666
Controls	No	Yes	No	Yes	No	Yes

Note: OLS regressions. The dependent variable is the first principal component in Panel A, and the negative of the second principal component in Panel B. Columns 1-2 report results for the full sample, Columns 3-4 report results for those with above median 2nd mock exam scores, and Columns 5-6 report results for those with below median 2nd mock exam scores. Scores are coded using percentile ranks. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

that STEM track choice increases with ability, as proxied by performance in the mock exam, and that at each ability level, boys are more likely to choose the STEM track than girls.

Our empirical approach is to first consider the established factors influencing STEM choice as identified in the literature, utilizing the data we have at our disposal, and then test whether response to competition has additional explanatory power. We estimate the following Probit model:

$$Pr(STEM_{ic} = 1) = \Phi(\alpha + \beta Female_{ic} + \Gamma K_{ic} + \Lambda D_{ic} + \Pi N_{ic} + \delta_c) \tag{2}$$

where $STEM_{ic} = 1$ if student i in cohort c chooses STEM track, and $= 0$ otherwise (i.e. chooses the non-STEM track). K_{ic} are a set of known determinants of the gender gap in

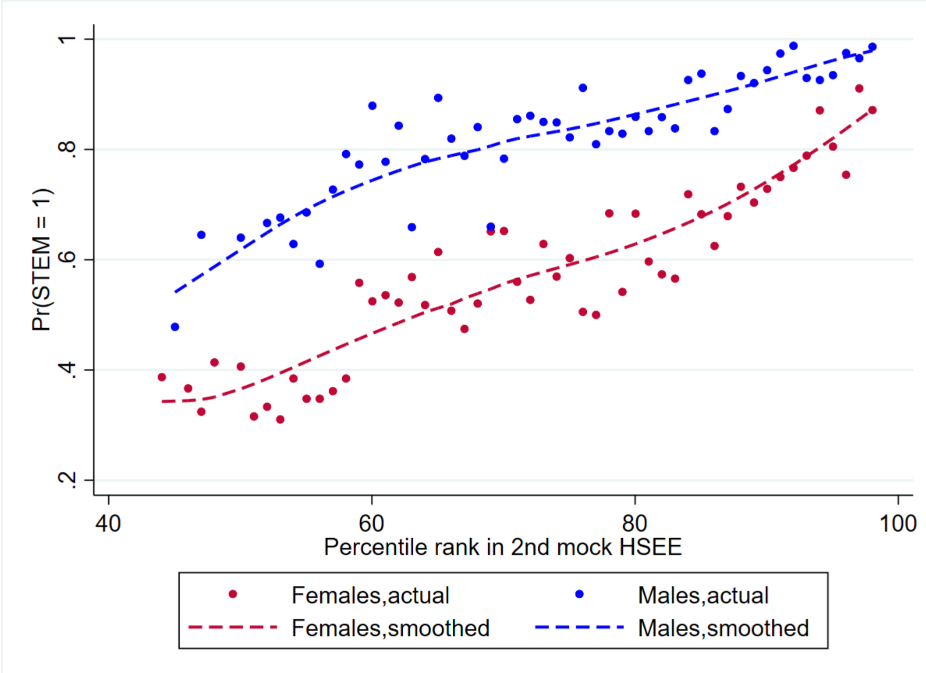


Figure 4: Probability of choosing STEM track by mock exam performance
Notes: Each dot represents the proportion choosing STEM track in a one-percentile point bin based on the percentile ranking in the 2nd mock exam. Bins with under 20 observations are omitted. Dashed lines are lowess smoothing lines.

choosing STEM. These include STEM ability, proxied by the student's STEM performance in the mock exam ($STEM^{Mock}$),[32] comparative advantage in STEM (CA^{Mock}), defined as the student's STEM performance relative to her non-STEM performance in the mock exam,[33] and peer or role model effects, proxied by the proportion of high performing female students in the student's high school class ($High\_Perf\_Female$).[34] D_{ic} are a set of background controls: age at HSEE, and SES, proxied by whether the individual graduated from an urban middle school and whether she comes from a poor family. N_{ic} are the response to competition variables introduced in this paper. We consider two such sets of

[32] Similar to our data, the literature finds that average STEM ability is similar for males and females (Hyde et al., 2008), but males are over represented in the tails (Ellison and Swanson, 2010).

[33] The literature finds that *relative* ability in STEM over non-STEM subjects has additional explanatory power for STEM choice, above and beyond STEM ability (Valla and Ceci, 2014; Breda and Napp, 2019; Goulas et al., 2022; Card and Payne, 2021; Loyalka et al., 2017).

[34] Following (Mouganie and Wang, 2020), this variable is defined as the proportion of females among the top 20% in each high school class. The findings on peer effects (see Mouganie and Wang (2020); Brenøe and Zölitz (2020); Bostwick and Weinberg (2022)) and on role model effects (see Carrell et al. (2010); de Gendre et al. (2023) have been mixed.

variables: response to competition in STEM (RC^{STEM}) and non-STEM ($RC^{non-STEM}$) , and domain-general (PC_1) and comparative advantage in response to competition (PC_2).

Results from estimating Equation 2 are reported in Table 8. Controlling only for cohort fixed effects in Column 1, the raw gender gap in choosing STEM is 24.2 percentage points, which is comparable in magnitude to the gender gap in choosing STEM or math intensive curricula in secondary school in other settings around the world.[35] [36] This equates to over one-third of the overall propensity to choose the STEM track. Controlling for STEM ability, comparative advantage in STEM, peer/role-model effects, as well as age and SES in Column (5) reduces the gender gap to 18.3 percentage points, which is a reduction of 24.4% of the raw gender gap in choosing STEM.

In Column (6) we introduce response to competition in STEM, finding it to be a statistically significant determinant of STEM choice. Furthermore, the gender gap in STEM choice reduces to 15.1 percentage points. Introducing response to competition in non-STEM in Column (7) further reduces the gender gap. The positive and significant coefficient on $RC^{non-STEM}$ indicates that higher response to competition in non-STEM is a deterrent from pursuing STEM studies, and that the female advantage in non-STEM response to competition enlarges the gender gap in STEM choice.

Gender differences in response to competition in STEM and non-STEM together contribute 4 percentage points to the gender gap in STEM choice. In other words, all else equal, if response to competition were not a factor in STEM choice, the gender gap in STEM choice

[35]For instance, the gender gap in taking Advanced Placement (AP) courses in math, computer science, and math-intensive sciences is 31.4 percentage points in the United States (Kahn and Ginther, 2017), and the raw gender gap in choosing the STEM track or the most math intensive track in secondary school is 34 percentage points in Greece (Goulas et al., 2022) and 23.4 percentage points in the Netherlands (Buser et al., 2014).

[36]When including the 2017 cohort, for which STEM choice is observed but data on responses to competition are missing, the observed gender gap in choosing STEM tracks stands at 24.6 percentage points.

would shrink by 4 percentage points. This represents 21.9% of the adjusted gender gap in column (5), and 16.5% of the raw gender gap in column (1).

Separate regressions for males and females are reported in Appendix Tables D.11 and D.12. Interestingly, for both males and females, the response to competition in STEM (RC^{STEM}) significantly explains STEM choice. However, only females' choices are significantly influenced by response to competition in non-STEM ($RC^{non-STEM}$), with no notable effect observed for males. Moreover, females exhibit a stronger reaction to the response to competition in both domains: the coefficient on RC^{STEM} ($RC^{non-STEM}$) is over twice (15 times) as large for females compared to males. These results potentially reflect greater beliefs updating in gender-congruent relative to gender-incongruent domains (Coffman et al., 2023). However, in this context where the outcome is choosing the STEM track, both males and females are more sensitive in absolute terms to RC^{STEM} than to $RC^{non-STEM}$.

We next test the extent to which domain-general and comparative advantage in response to competition explains STEM choice and the gender gap in STEM choice. This is an open question in the literature, although the dominant view favors domain-generality. In Table 9 we reproduce the first five columns of Table 8 and additionally include the domain-general and comparative advantage components of response to competition as defined in Section 5.5 in Columns (6) and (7).

We find in Column (6) that the domain-general component of response to competition (i.e., the first principal component) is a positive and significant determinant of STEM choice. This supports the dominant view that STEM is a competitive field, and that those who respond better to competition across domains are more likely to choose STEM (see Buser et al. (2014, 2017, 2024) on explaining STEM choice using lab elicited and self-reports of willingness to compete). However, the domain-general component does not diminish the

Table 8: Determinants of STEM-track choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var = 1 if STEM track							
Female	-0.242***	-0.208***	-0.177***	-0.178***	-0.183***	-0.151***	-0.143***
	(0.011)	(0.012)	(0.013)	(0.014)	(0.014)	(0.014)	(0.015)
$STEM^{Mock}$		1.259***	1.199***	0.942***	1.067***	1.024***	1.030***
		(0.038)	(0.039)	(0.047)	(0.070)	(0.079)	(0.077)
CA^{Mock}			0.313***	0.392***	0.333***	0.442***	0.465***
			(0.053)	(0.062)	(0.064)	(0.066)	(0.067)
High_Perf_Females				-0.005***	-0.006***	-0.006***	-0.006***
				(0.000)	(0.000)	(0.000)	(0.000)
RC^{STEM}						1.402***	1.453***
						(0.120)	(0.118)
$RC^{non-STEM}$							-0.249**
							(0.107)
Controls	No	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.052	0.236	0.240	0.250	0.254	0.284	0.285
Observations	6,452	6,452	6,452	5,071	5,071	5,071	5,071

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

gender gap in STEM choice. On the other hand, we find in Column (7) that the comparative advantage component of response to competition (i.e., the negative of the second principal component) is both a significant determinant of STEM choice and contributes to the gender gap in STEM choice. It alone accounts for 3.4 percentage points of the gap. The coefficient is negative, as expected, which means that, again, girls' advantage in response to competition in non-STEM over STEM enlarges the gender gap in STEM choice. Appendix D.13 and D.14 reproduces the above analysis separately for the elite and regular high school students. The patterns are the same as that found in Table 9, with both components explaining STEM choice, and the comparative advantage component explaining a larger fraction of the gender gap in STEM choice than the domain-general component.

Table 9: Determinants of STEM-track choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var = 1 if STEM track							
Female	-0.242***	-0.208***	-0.177***	-0.178***	-0.183***	-0.183***	-0.149***
	(0.011)	(0.012)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)
$STEM^{Mock}$		1.259***	1.199***	0.942***	1.067***	0.947***	1.009***
		(0.038)	(0.039)	(0.047)	(0.070)	(0.073)	(0.075)
CA^{Mock}			0.313***	0.392***	0.333***	0.400***	0.410***
			(0.053)	(0.062)	(0.064)	(0.065)	(0.066)
High_Perf_Females				-0.005***	-0.006***	-0.006***	-0.006***
				(0.000)	(0.000)	(0.000)	(0.000)
PC_1						0.051***	0.058***
						(0.008)	(0.008)
$-PC_2$							-0.086***
							(0.008)
Controls	No	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.052	0.236	0.240	0.250	0.254	0.264	0.284
Observations	6,452	6,452	6,452	5,071	5,071	5,071	5,071

Notes: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. PC_1 and PC_2 are the first two principal components in a PCA of response to competition in all seven subjects in the HSEE, as described in Section 5.5. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Separate regressions for males and females are reported in Appendix Tables D.15 and D.16. Similar to the full sample analysis, in each sub-sample, both the domain-general and comparative advantage components significantly explain STEM choice. Moreover, the impact on STEM choice is more pronounced among females; the marginal effects of PC_1

and PC_2 for females is more than double that for males. This aligns with the above analysis by gender, where we found girls to be more sensitive to response to competition in both domains. The relative insensitivity of males to these determinants of STEM choice could arise from the importance of the STEM wage premium for males. Zafar (2013) finds that non-pecuniary factors play a larger role in female major choice, while pecuniary factors are much more important for males.

7 Stability and persistence

The previous section suggests that gender gaps in response to competition measured pre-market could impact long-term labor market outcomes through irreversible track choices. Another potential channel is the stability and persistence of gender gaps in response to competition over time. In this section we check for the temporal stability and persistence of response to competition using the College Entrance Exam (CEE), which is taken 3 years after the HSEE. The CEE, like the HSEE, is a highly competitive exam and performance in the CEE is the sole determinant of priority in the college admissions process (see Cai et al. (2019)). Four non-competitive but informative mock exams are held prior to the CEE. Due to the fact that high school students in STEM and non-STEM tracks follow different curricula, we focus the analysis on the subjects common to all students - Math, Chinese Language, and English Language. Response to competition in the CEE is defined as the (residualized) difference between performance in the CEE and the 4th mock CEE, similar to how it is defined in the HSEE.

7.1 Stability over time

Temporal stability in response to competition implies that those who respond well to competition at one point in time will also respond well at a later time. In Table 10, we report correlation coefficients of (residualized) response to competition in the CEE and (residualized) response to competition in the HSEE. The results show that for all three subjects common to the STEM and non-STEM tracks, response to competition in the HSEE is highly correlated with that in the CEE. Controlling for age and SES, proxied by poverty status and whether the individual attended an urban middle school, has virtually no impact on the strength of the relationship. The correlation coefficients of 0.18-0.20 (using percentile ranks) in Math and Chinese indicate that the temporal stability of response to competition in these subjects are similar to that found for risk preferences over a similar time frame (see Schildberg-Hörisch (2018) and Chuang and Schechter (2015)). Appendix Table D.17 reproduces this table separately for males and females, finding very similar correlation coefficients across gender.

7.2 Persistence

Although the above analysis revealed significant evidence of temporal stability, the stability is not perfect, and does not necessarily imply that the same gender patterns of response to competition will persist over time. To check for persistence, we estimate Equation 1 but replace the outcome variable with response to competition in the CEE: the difference in performance between the CEE and the last mock CEE. The results are reported in Table 11.

Two findings emerge from this exercise. First, the pattern of domain-specific gender gaps in response to competition still exists in the CEE: specifically, women have lower response to competition than men in the STEM subject (Math), but higher response to competition in the Non-STEM subjects (Chinese and English).[37] Second, the magnitude of the gender gap in response to competition is more pronounced in the CEE compared to the HSEE. In the HSEE, the female disadvantage in STEM, for instance, was 1 percentile or

[37] In Appendix table D.18 we replicate Table 11 but leave out the control for performance in the mock exam, and find that the female coefficient is negative for both STEM and non-STEM subjects, which reiterates the importance of omitted variable bias when not accounting for the mechanical relationship between response to competition and mock exam performance.

Table 10: Correlation between response to competition in the HSEE and the CEE

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$RC^{Math-CEE}$		$RC^{Chinese-CEE}$		$RC^{English-CEE}$	
Panel A: Percentile ranking						
$RC^{Math-HSEE}$	0.177***	0.178***				
	(0.020)	(0.020)				
$RC^{Chinese-HSEE}$			0.198***	0.198***		
			(0.017)	(0.017)		
$RC^{English-HSEE}$					0.107***	0.105***
					(0.020)	(0.020)
Panel B: Standardized score						
$RC^{Math-HSEE}$	0.136***	0.134***				
	(0.019)	(0.019)				
$RC^{Chinese-HSEE}$			0.182***	0.183***		
			(0.018)	(0.018)		
$RC^{English-HSEE}$					0.068***	0.065***
					(0.019)	(0.019)
Controls	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Track	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,760	3,760	3,760	3,760	3,760	3,760

Notes: Correlation coefficients between residualized response to competition in the HSEE and in the CEE. Scores are coded using percentile ranks in Panel A and using standarized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

3.5 percent of a standard deviation, while in the CEE it is 2.7 percentiles or 7.8 percent of a standard deviation. These findings are consistent with research showing that once gender differences in competitiveness emerge around the age of 15, they persist into adulthood (Andersen et al., 2013), and resonate with the fact that the under-representation of women in STEM intensifies at each successive education and career stage (Kahn and Ginther, 2017). Appendix Table D.19 reproduces the above analysis separately for STEM and non-STEM track students and find similar gender coefficients across the two tracks.

The findings of this and the previous section suggest two potential mechanisms by which pre-market measures of response to competition evolve into gender inequality in the labor market: First, due to gender differences in domain-specific response to competition, girls are less likely to choose the STEM track, which prevents them from choosing the more lucrative STEM majors in college. Second, compared with boys, girls consistently underperform in competitions in the STEM domain and overperform in competitions in the non-STEM domain over time, relative to their abilities, and these patterns intensify over time. Therefore, at the same ability level, conditional on having selected the STEM track, girls may opt-out or drop out of STEM majors and STEM careers with greater frequency than boys. The second mechanism will apply more generally to education systems in which track or field choice is more flexible or is made at a later stage.

8 Conclusion and discussion

Competitions are used to select those most deserving of limited slots in education and the labor market. This practice is not only ubiquitous but embraced as a hallmark of meritocracy. Competitive selection processes, however, may select not only the most able, but also those who perform particularly well under competition. These may not be the same groups of people.

The findings of this paper show that girls are disadvantaged by their lower response to competition in STEM in entry into high school (and advantaged by their higher response to competition in non-STEM), that girls' lower response to competition in STEM relative to non-STEM results in discouraging them from choosing the STEM track, which has long term consequences for the college majors they are eligible to apply for, and that girls' disadvantage in response to competition in STEM widens by the time of college applications. To the extent that response to competition do not reliably enhance labor market productivity, gender differences in response to competition are a potential source of job and occupational mismatch, which adversely impacts both genders.

Relatedly, the debate on the use of standardized tests in the college admissions process has focused on their predictive value for college GPA and potential socioeconomic and racial bias (Miller et al., 2014; Douglass, 2015; Wai et al., 2018; Duquennois, 2022). Our results suggest that in addition to these potential shortcomings, students themselves may use their performance shocks in high stakes competitive exams to draw conclusions about their aptitude for different academic tracks in a way that exaggerates and perpetuates gender gaps in STEM choice above and beyond that predicted by their domain-specific ability.

A potential policy intervention addressing these findings could be to reduce the role of competitiveness in measuring performance. For example, once-and-for-all exams could be replaced with multiple exams. Suggestively, Cotton et al. (2013) find a male advantage in the first of a series of math contests, but none in subsequent ones, and none when competitive pressure is removed. Beijing is trialling a policy of allowing students to take the CEE twice and use the highest score to determine college admissions, rather than the single CEE.[38] We leave to future research to examine the impact of such a change in the educational institution.

[38]See Beijing's *2014-2016 National College Entrance Examination (Gaokao) and Higher Education Admission Reform Framework Plan*: http://politics.people.com.cn/BIG5/n/2013/1021/c1001-23277975.html

Table 11: Response to Competition (CEE)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Math		Chinese		English	
Panel A: Percentile rank						
Female	-2.037***	-2.556***	2.672***	3.852***	1.072**	1.577***
	(0.583)	(0.579)	(0.808)	(0.750)	(0.444)	(0.448)
$Math^{Mock}$	-0.223***	-0.302***				
	(0.009)	(0.013)				
$Chinese^{Mock}$			-0.450***	-0.613***		
			(0.013)	(0.015)		
$English^{Mock}$					-0.125***	-0.170***
					(0.007)	(0.010)
Constant	14.302***	33.935***	24.493***	80.420***	8.489***	16.270***
	(0.916)	(6.599)	(1.160)	(8.004)	(0.656)	(4.755)
Controls	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,764	3,764	3,764	3,764	3,764	3,764
R-squared	0.118	0.144	0.224	0.326	0.072	0.088
Panel B: Standardized score						
Female	-0.060***	-0.078***	0.092***	0.121***	0.040***	0.061***
	(0.019)	(0.019)	(0.028)	(0.026)	(0.014)	(0.015)
$Math^{Mock}$	-0.232***	-0.310***				
	(0.010)	(0.013)				
$Chinese^{Mock}$			-0.444***	-0.596***		
			(0.030)	(0.032)		
$English^{Mock}$					-0.109***	-0.164***
					(0.009)	(0.012)
Constant	0.114***	0.677***	0.055*	1.623***	0.057***	0.248
	(0.024)	(0.214)	(0.031)	(0.254)	(0.017)	(0.159)
Controls	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,764	3,764	3,764	3,764	3,764	3,764
R-squared	0.134	0.159	0.208	0.316	0.055	0.075

Note: OLS regressions. The dependent variable is response to competition in the CEE. Columns 1-2 report results for Math, Columns 3-4 report results for Chinese Language, and Columns 5-6 report results for English Language. *Mock* refers to scores on the 4th and last mock CEE. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an elite high school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

References

Abdulkadiroğlu, A., Che, Y.-K., and Yasuda, Y. (2011). Resolving conflicting preferences in school choice: The" boston mechanism" reconsidered. *American Economic Review*, 101(1):399–410.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ.*, 4(1):185–223.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., and 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., and Iriberri, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6):1372–1400.

Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., and Doms, M. E. (2011). Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11).

Bertrand, M. (2011). New perspectives on gender. In *Handbook of labor economics*, volume 4, pages 1543–1590. Elsevier.

Bietenbeck, J. and Collins, M. (2023). New evidence on the importance of instruction time for student achievement on international assessments. *Journal of Applied Econometrics*, 38(1):145–166.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059.

Bostwick, V. K. and Weinberg, B. A. (2022). Nevertheless she persisted? gender peer effects in doctoral stem programs. *Journal of Labor Economics*, 40(2):397–436.

Breda, T. and Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences*, 116(31):15435–15440.

Brenøe, A. A. and Zölitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4):1009–1054.

Buser, T. (2016). The impact of losing in a competition on the willingness to seek further challenges. *Management Science*, 62(12):3439–3449.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The quarterly journal of economics*, 129(3):1409–1447.

Buser, T., Niederle, M., and Oosterbeek, H. (2024). Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures. *Review of Economics and Statistics*, pages 1–45.

Buser, T., Peter, N., and Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American economic review*, 107(5):125–130.

Cai, X., Lu, Y., Pan, J., and Zhong, S. (2019). Gender gap under pressure: Evidence from china's national college entrance examination. *Review of Economics and Statistics*, 101(2):249–263.

Card, D. and Payne, A. A. (2021). High school choices and the gender gap in stem. *Economic Inquiry*, 59(1):9–28.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Carnevale, A. P., Smith, N., and Melton, M. (2011). Stem: Science technology engineering mathematics. *Georgetown University Center on Education and the Workforce*.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly journal of economics*, 125(3):1101–1144.

Cassar, A. and Rigdon, M. (2023). Sustaining the potential for cooperation as female competitive strategy. *Philosophical Transactions of the Royal Society B*, 378(1868):20210440.

Cassar, A. and Rigdon, M. L. (2021a). Option to cooperate increases women's competitiveness and closes the gender gap. *Evolution and Human Behavior*, 42(6):556–572.

Cassar, A. and Rigdon, M. L. (2021b). Prosocial option increases women's entry into competition. *Proceedings of the National Academy of Sciences*, 118(45):e2111943118.

Cassar, A., Wordofa, F., and Zhang, Y. J. (2016). Competing for the benefit of offspring eliminates the gender gap in competitiveness. *Proceedings of the National Academy of Sciences*, 113(19):5201–5205.

Cassar, A. and Zhang, Y. J. (2022). The competitive woman: Evolutionary insights and cross-cultural evidence into finding the femina economica. *Journal of Economic Behavior & Organization*, 197:447–471.

Chuang, Y. and Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of development economics*, 117:151–170.

Coffman, K., Collis, M. R., and Kulkarni, L. (2023). Stereotypes and belief updating. *Journal of the European Economic Association*, page jvad063.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.

Cotton, C., McIntyre, F., and Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, 86:52–66.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–474.

de Gendre, A., Feld, J., Salamanca, N., and Zölitz, U. (2023). Same-sex role model effects in education. *Working paper series/Department of Economics*, (438).

Dee, T. S. (2014). Stereotype threat and the student-athlete. *Economic Inquiry*, 52(1):173–182.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3):522–550.

Douglass, J. A. (2015). Sat wars at the university of california.”. *SAT Wars: The Case for Test-Optional College Admissions*, pages 50–68.

Duquennois, C. (2022). Fictional money, real costs: Impacts of financial salience on disadvantaged students. *American Economic Review*, 112(3):798–826.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2017). The long-term effects of early track choice. *The Economic Journal*, 127(603):1348–1380.

Ellison, G. and Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the american mathematics competitions. *Journal of Economic Perspectives*, 24(2):109–128.

Ergin, H. and Sönmez, T. (2006). Games of school choice under the boston mechanism. *Journal of public Economics*, 90(1-2):215–237.

Francesconi, M. and Parey, M. (2018). Early gender gaps among university graduates. *European economic review*, 109:63–82.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., and Xu, Y. (2019). Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights*, 1(3):291–308.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The quarterly journal of economics*, 118(3):1049–1074.

Goulas, S., Griselda, S., and Megalokonomou, R. (2022). Comparative advantage and gender gap in stem. *Journal of Human Resources*, pages 0320–10781R2.

Günther, C., Ekinci, N. A., Schwieren, C., and Strobel, M. (2010). Women can't jump?—an experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3):395–401.

Hu, A. and Vargas, N. (2015). Economic consequences of horizontal stratification in postsecondary education: Evidence from urban china. *Higher Education*, 70:337–358.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, 321(5888):494–495.

Iriberri, N. and 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.

Kahn, S. and Ginther, D. (2017). Women and science, technology, engineering, and mathematics (stem): Are differences in education and careers due to stereotypes, interests, or family?

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? evidence from a natural experiment. *Journal of public Economics*, 92(10-11):2083–2105.

Lavy, V. and Megalokonomou, R. (2019). Persistency in teachers' grading bias and effects on longer-term outcomes: University admissions exams and choice of field of study. Technical report, National Bureau of Economic Research.

Lavy, V. and Sand, E. (2018). On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics*, 167:263–279.

Li, Q., Liu, G., and Zang, W. (2015). The health of left-behind children in rural china. *China Economic Review*, 36:367–376.

Li, S. (2007). Imbalanced sex ratio at birth and comprehensive intervention in china.

Liu, M., Hu, W., Jiannong, S., and Adey, P. (2010). Gender stereotyping and affective attitudes towards science in chinese secondary school students. *International Journal of Science Education*, 32(3):379–395.

Loyalka, P. K., Maani, M., Qu, Y., and Sylvia, S. (2017). Absolute versus comparative advantage: Consequences for gender gaps in stem and college access. *Available at SSRN 2908533*.

Miller, C., Stassun, K., et al. (2014). A test that fails. *Nature*, 510(7504):303–304.

Morin, L.-P. (2015). Do men and women respond differently to competition? evidence from a major education reform. *Journal of Labor Economics*, 33(2):443–491.

Mouganie, P. and Wang, Y. (2020). High-performing peers and female stem choices in school. *Journal of Labor Economics*, 38(3):805–841.

Murphy, R. and Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, 87(6):2777–2826.

Niederle, M. and 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. and Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of economic perspectives*, 24(2):129–44.

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European review of social psychology*, 18(1):36–88.

Ors, E., Palomino, F., and Peyrache, E. (2013). Performance gender gap: does competition matter? *Journal of Labor Economics*, 31(3):443–499.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29(6):923–934.

Peng, X. (2011). China's demographic history and future challenges. *science*, 333(6042):581–587.

Rask, K. and Tiefenthaler, J. (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, 27(6):676–687.

Reuben, E., Sapienza, P., Zingales, L., et al. (2019). Taste for competition and the gender gap among young business professionals. *REPEC Working Papers 20190031*.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Roberts, T.-A. and Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback. *Sex Roles*, 21:725–747.

Sancassani, P. (2022). The effect of teacher subject-specific qualifications on student science achievement. *Labour Economics*, 72:102309.

Santos-Pinto, L. and Sekeris, P. (2023). Confidence and gender gaps in competitive environments. *Available at SSRN 4479952*.

Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, 32(2):135–154.

Schlosser, A., Neeman, Z., and Attali, Y. (2019). Differential performance in high versus low stakes tests: Evidence from the gre test. *The Economic Journal*, 129(623):2916–2948.

Schmader, T., Johns, M., and Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological review*, 115(2):336.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5):1189–1213.

Shurchkov, O. and Eckel, C. C. (2018). Gender differences in behavioral traits and labor market outcomes.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, 35(1):4–28.

Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology*, 69(5):797.

Terrier, C. (2020). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review*, 77:101981.

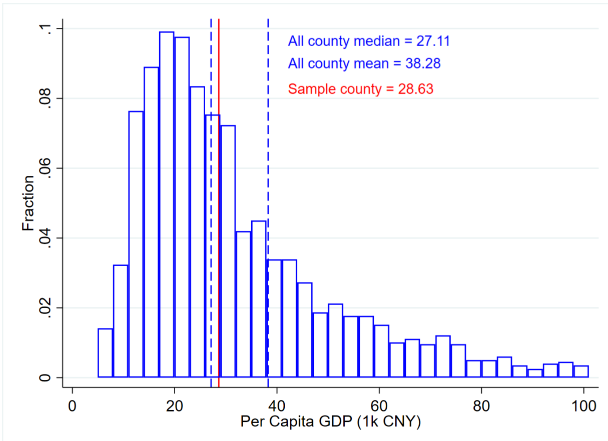
Valla, J. M. and Ceci, S. J. (2014). Breadth-based models of women's underrepresentation in stem fields: An integrative commentary on schmidt (2011) and nye et al.(2012). *Perspectives on Psychological Science*, 9(2):219–224.

Wai, J., Brown, M. I., and Chabris, C. F. (2018). Using standardized test scores to include general cognitive ability in education research and policy. *Journal of Intelligence*, 6(3):37.

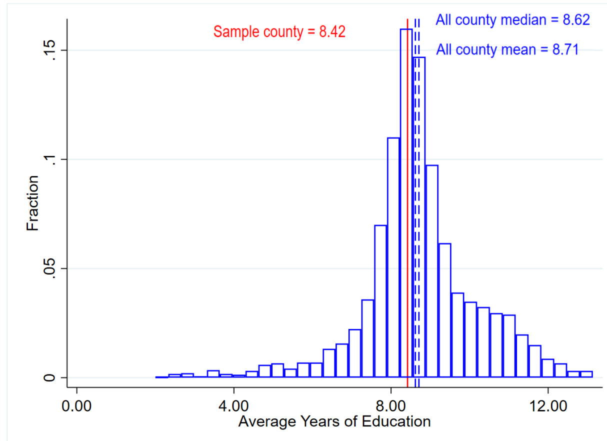
Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

Zhang, Y. J. (2019). Culture, institutions and the gender gap in competitive inclination: Evidence from the communist experiment in china. *The Economic Journal*, 129(617):509–552.

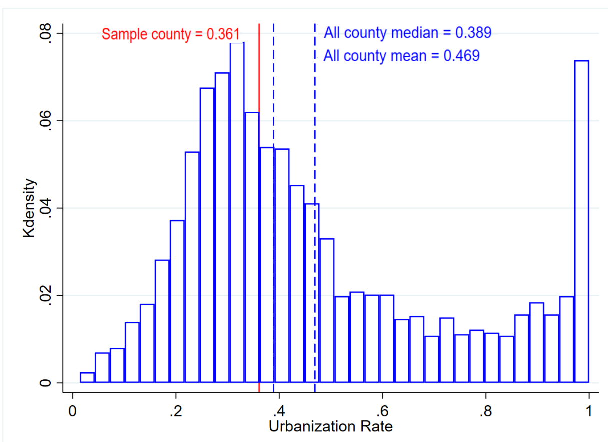
A Appendix A: Representativeness of the sample county



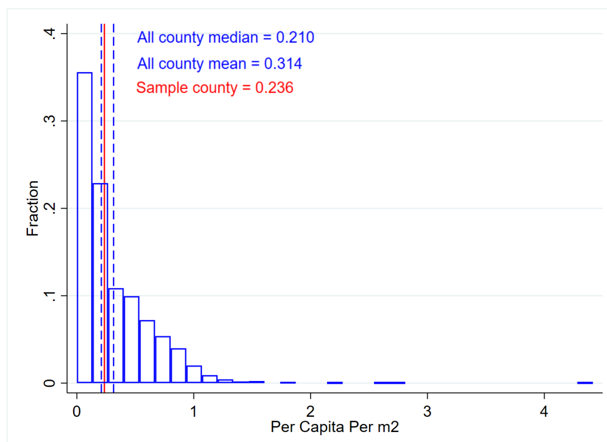
(a) Per Capita GDP(1 K CNY)



(b) Average Years of Education



(c) Urbanization Rate



(d) Population Density

Figure A1: Sample County Characteristics

B Appendix B: Data merging and cleaning process

B.1 Procedure for merging different datasets

This appendix outlines the data merging process employed in this study. Since both the HSEE and CEE datasets utilize the national ID as a unique identifier, they can be directly linked. It is from the CEE data that we extract the track choice for all students who proceed to academic high schools. The challenge arises in linking the HSEE data with HSEE mock exams and the CEE data with CEE mock exams, due to the absence of national ID information in the mock exam datasets. We describe the two merging procedures below:

B.1.1 Merge HSEE mock exam scores with HSEE scores

- STEP 1: For each mock exam, we first keep those students who can be uniquely identified by name, school name, classroom and exam year. 0.16% are dropped in this step.

- STEP 2: For those with classroom information, we use their name, school name, class number and exam year to match with students from the HSEE data who can be uniquely identified by this information. Of the overall sample, 85.59% is matched in this way.

- STEP 3: For those left from step 2, we employ the Pinyin spelling of names along with additional information to re-match with the HSEE score data that remained unmatched from step 2. This approach accounts for the presence of homophones with different characters in student names. Through this method, 0.59% of the total sample is successfully matched.

- STEP 4: For those left from step 3, we keep those who can be uniquely identified by their name, school name and exam year, and merge with the unmatched HSEE score data from step 3. The purpose of this step is to match students with classroom information missing in the mock exam data. 11.4% of the overall sample is matched in this way.

- STEP 5: For those unmatched by the previous steps, two RAs attempted to match records manually, and we kept the matches that were agreed on by both RAs. The purpose of this step is to deal with student names that have been written incorrectly by accident. For instance, Wang Shuangshuang might be written incorrectly as Wang Shuang. 0.82% are matched in this way.

B.1.2 Merge 4th CEE mock exam scores and CEE scores

We follow a similar process as with the HSEE:

- STEP 1: We first keep those students who can be uniquely identified by school name, exam year, track and student name in both dataset. 1.3% are dropped in this step.
- STEP 2: We then use school name + cohort + track + student name to match with students from CEE data set and also can be uniquely identified by those information. 92.15% of the overall sample can be matched in this way.
- STEP 3: For those left from step 2, we keep those that can be uniquely identified by cohort + track + student name (school information missing), and merge again. This step takes into consideration those with missing school information in the mock exam. 5.19% of the overall sample is matched in this way.
- STEP 4: For those left from step 2, we keep those that can be uniquely identified by cohort + school name + student name (track information missing), and merge again. This step takes into consideration those with missing track information in the mock exam. 0.2% of the overall sample is matched in this way.
- STEP 5: For those left from step 3 and 4, we use the name pinyin spelling and other information to match again. This step is to take into consideration of the presence of

homophones with different characters in student names. 0.45% of the total sample is matched in this way.

B.2 Additional statistics

Table B.1: Descriptive Statistics for all available data

Explanation	obs	Mean	Std
Overall sample			
1. Middle school			
Years available: 2015, 2016, 2018, 2019			
Total score in HSEE	20,100	324.8	111.7
STEM subjects score in HSEE	20,100	117.7	58.3
Non-STEM subjects score in HSEE	20,100	207.1	57.3
Total score in 2nd mock	19,686	319.2	103.2
STEM subjects score in 2nd mock	19,686	116.0	52.2
Non-STEM subjects score in 2nd mock	19,686	203.2	55.1
Female dummy	20,186	0.46	0.50
Age at HSEE	20,186	15.60	0.70
Urban middle school dummy	20,186	0.61	0.49
Poverty status dummy	20,186	0.13	0.34
2. High school			
Years available: 2015-2018			
STEM dummy	8,698	0.66	0.47
Elite high school dummy	9,004	0.46	0.50
Female dummy	9,004	0.50	0.50
Age at HSEE	9,004	185.57	8.18
Urban middle school dummy	9,004	0.77	0.42
Poverty status dummy	9,004	0.10	0.30
3. College Entrance Exam			
Years available: 2018-2019			
Total score in CEE	3,854	420.2	97.6
STEM subjects score in CEE	3,854	248.6	80.6
Non-STEM subjects score in CEE	3,854	171.6	104.1
Total score in 4th mock	3,798	402.0	95.6
STEM subjects score in 4th mock	3,798	204.4	79.0
Non-STEM subjects score in 4th mock	3,798	153.6	102.9
Elite high school dummy	3,854	0.52	0.50
Female dummy	3,854	0.53	0.50
age at CEE	3,854	18.49	0.69
Poverty status dummy	3,854	0.10	0.31

Table B.2: Proportion taking the HSEE and high school admissions rate

Year	Middle School Graduating Cohort	HSEE taking rate	High school admissions rate	Elite high school admissions rate
2015	5,274	89.25	52.48	23.48
2016	5,429	87.51	48.60	22.29
2017	5,538	93.66	46.93	20.32
2018	5,638	94.18	46.18	19.83
2019	5,634	94.78	47.15	20.60

C Appendix C: Admission procedures in Chinese high schools

Table C.1: HSEE Schedule

Date	Time	Duration	Subject	Full Mark
Day 1	8:30AM - 10:30AM	120 mins	Chinese	120
	11:10AM - 12:00AM	50 mins	History	50
	3:00PM - 4:00PM	70 mins	Physics	70
	4:40PM - 5:30Pm	50 mins	Chemistry	50
Day 2	8:30AM - 10:10AM	100 mins	Maths	120
	10:50AM - 11:50AM	60 mins	Politics	70
	3:00PM - 4:40PM	100 mins	English	120

The provincial education bureau first determines each high school's student capacity for the coming year. Typically, including in the county where we collect data, high schools can only admit students within the county. In the first round, the education bureau considers the first-choice high schools of all students within the county, and admits students until capacity is reached at each high school. HSEE scores are used to determine priority throughout the process. In the second round, after admitted students are removed from consideration, the education bureau admits students based on their second-choices, provided the high schools have remaining capacity after the first round. The process continues until all high schools have reached capacity (or there are no more students to be admitted). As is well-known of the Boston mechanism, once a student is admitted, they are no longer considered in any subsequent rounds, in contrast to the deferred acceptance mechanism, see, for example Ergin and Sönmez (2006) and Abdulkadiroğlu et al. (2011). This means that if a marginal student whose HSEE score is high enough for admission into regular high school indicates an elite high school as their first choice, they may not be admitted in the first round, and may miss out on being admitted to regular high schools as well, if regular high schools are fully subscribed after the first round. Responding strategically to this choice situation, a student may put down a regular high school as their first choice and be admitted, when they would have been admitted to an elite high school, presumably their true first choice, had that been their indicated first choice.

With two high schools in the county and presumably uniform true preferences between the two high schools (i.e., the elite high school is preferred over the regular high school), the process can be improved by a centralized one where no list of preference is sought and admissions are determined by HSEE scores. The reason ranked ordered lists are used is that admissions is administered by the provincial education bureau, and the typical county in this province has multiple elite and regular high schools where preferences are not uniform between high schools in the same category.

D Appendix D: Other empirical results

Table D.1: Response to competition (HSEE) - controlling class size

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM subjects			Non-STEM subjects		
Female	-0.902***	-0.906***	-0.977***	-0.321**	0.387***	0.437***
	(0.152)	(0.150)	(0.150)	(0.139)	(0.143)	(0.143)
$Mock^{\text{STEM}}$		-1.919***	-2.024***			
		(0.060)	(0.071)			
$Mock^{\text{Non-STEM}}$					-1.760***	-1.926***
					(0.061)	(0.070)
Constant	0.757***	0.759***	17.677***	0.503***	0.164	11.437***
	(0.174)	(0.171)	(1.896)	(0.163)	(0.160)	(1.751)
Observations	19,549	19,549	19,458	19,549	19,549	19,458
R-squared	0.002	0.035	0.041	0.001	0.032	0.035

Note: OLS regressions. The dependent variable is response to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the second mock exam. Scores are coded using percentile ranks. Controls include age at HSEE, a dummy for graduating from an urban middle school, a dummy for being from a poor family, and middle school class size. Robust standard errors are in parentheses.

Table D.2: Response to competition (HSEE) - controlling within-class rank

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM subjects			Non-STEM subjects		
Female	-0.902***	-0.906***	-0.971***	-0.321**	0.387***	0.508***
	(0.152)	(0.150)	(0.150)	(0.139)	(0.143)	(0.144)
$Mock^{\text{STEM}}$		-1.919***	-1.838***			
		(0.060)	(0.135)			
$Mock^{\text{Non-STEM}}$					-1.760***	-1.543***
					(0.061)	(0.128)
Constant	0.757***	0.759***	14.654***	0.503***	0.164	11.467***
	(0.174)	(0.171)	(1.886)	(0.163)	(0.160)	(1.751)
Observations	19,549	19,549	19,458	19,549	19,549	19,458
R-squared	0.002	0.035	0.038	0.001	0.032	0.035

Note: OLS regressions. The dependent variable is response to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the second mock exam. Scores are coded using percentile ranks. Controls include age at HSEE, a dummy for graduating from an urban middle school, a dummy for being from a poor family and within middle school class percentile ranks in the second mock exam. Robust standard errors are in parentheses.

Table D.3: Response to competition (HSEE) – controlling interaction terms

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM subjects			Non-STEM subjects		
Marginal effect	-0.902***	-0.906***	-0.973***	-0.321**	0.339**	0.247**
	(0.152)	(0.150)	(0.163)	(0.139)	(0.144)	(0.156)
Interaction	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19,549	19,549	19,549	19,549	19,549	19,549
R-squared	0.002	0.035	0.035	0.001	0.033	0.033

Note: OLS regressions. The dependent variable is response to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. All specifications include a dummy for female and scores on the second mock exam in the relevant domain. Scores are coded using percentile ranks. Interaction is the interaction between the female dummy and scores on the second mock exam. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.4: Response to competition (HSEE) - High-performing subsample

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM subjects			Non-STEM subjects		
Panel A: Percentile ranking						
Female	-0.492***	-0.949***	-1.104***	0.865***	1.174***	1.057***
	(0.183)	(0.186)	(0.185)	(0.178)	(0.181)	(0.181)
Mock$^{\text{STEM}}$		-1.767***	-1.653***			
		(0.143)	(0.154)			
Mock$^{\text{Non-STEM}}$					-1.886***	-1.832***
					(0.171)	(0.173)
Constant	-0.372*	1.333***	16.163***	-1.017***	0.343	8.247***
	(0.193)	(0.258)	(2.325)	(0.208)	(0.259)	(2.303)
Controls	No	No	Yes	No	No	Yes
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,883	9,883	9,883	9,883	9,883	9,883
R-squared	0.002	0.015	0.023	0.004	0.017	0.020
Panel B: Standardized score						
Female	0.011*	-0.035***	-0.041***	0.037***	0.044***	0.040***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Mock$^{\text{STEM}}$		-0.180***	-0.176***			
		(0.005)	(0.006)			
Mock$^{\text{Non-STEM}}$					-0.042***	-0.040***
					(0.006)	(0.006)
Constant	-0.019***	0.155***	0.659***	-0.034***	-0.003	0.269***
	(0.007)	(0.009)	(0.083)	(0.007)	(0.009)	(0.079)
Controls	No	No	Yes	No	No	Yes
Year dummy	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,883	9,883	9,883	9,883	9,883	9,883
R-squared	0.001	0.099	0.106	0.005	0.011	0.014

Note: OLS regressions. The sample consists of the top 50th percentile based on scores in the second mock exam. The dependent variable is respones to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the second mock exam. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.5: Performance improvement between the mock exams – High-performing subsample

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM subjects			Non-STEM subjects		
Panel A: Percentile ranking						
Female	-0.517**	0.295	0.212	0.403*	-0.135	-0.142
	(0.230)	(0.232)	(0.247)	(0.240)	(0.236)	(0.249)
					(0.171)	(0.173)
*Mock*1	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observations	4,262	4,262	4,262	4,262	4,262	4,262
R-squared	0.181	0.225	0.235	0.151	0.184	0.187
Panel B: Standardized score						
Female	-0.007	0.013	0.011	0.008	-0.003	-0.003
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)
*Mock*1	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observations	4,262	4,262	4,262	4,262	4,262	4,262
R-squared	0.015	0.037	0.052	0.009	0.022	0.023

Note: OLS regressions using the top 50th percentile sample based on scores in the second mock exam. The dependent variable is the 2nd mock exam score minus the 1st mock exam score. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock*1 refers to scores on the first mock exam. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.6: Response to competition (HSEE) - 2018-2019 subsample

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES		STEM subjects			Non-STEM subjects	
Panel A: Percentile rank						
Female	-0.941***	-0.953***	-1.063***	-0.703***	-0.026	0.003
	(0.193)	(0.191)	(0.190)	(0.184)	(0.190)	(0.191)
$Mock^{\text{STEM}}$		-1.676***	-1.510***			
		(0.077)	(0.092)			
$Mock^{\text{Non-STEM}}$					-1.625***	-1.769***
					(0.088)	(0.102)
Constant	0.548***	0.556***	13.884***	0.466***	0.159	10.910***
	(0.164)	(0.162)	(2.367)	(0.154)	(0.151)	(2.409)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,313	10,313	10,313	10,313	10,313	10,313
R-squared	0.002	0.032	0.039	0.001	0.030	0.033
Panel B: Standardized score						
Female	-0.018***	-0.019***	-0.022***	-0.025***	0.002	0.004
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$Mock^{\text{STEM}}$		-0.064***	-0.061***			
		(0.003)	(0.003)			
$Mock^{\text{Non-STEM}}$					-0.063***	-0.072***
					(0.005)	(0.005)
Constant	0.015**	0.015***	0.492***	0.019***	0.007	0.430***
	(0.006)	(0.006)	(0.085)	(0.006)	(0.005)	(0.087)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,313	10,313	10,313	10,313	10,313	10,313
R-squared	0.001	0.034	0.039	0.001	0.034	0.039

Note: OLS regressions. The sample consists of the 2018-2019 cohorts, i.e., the cohorts for whom the 1st mock exam data are available. The dependent variable is response to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the second mock exam. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.7: Response to competition (HSEE) - 2018-2019 High-Performing subsample

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	STEM subjects			Non-STEM subjects		
Panel A: Percentile rank						
Female	-0.412*	-0.931***	-1.067***	0.723***	1.152***	1.068***
	(0.234)	(0.240)	(0.240)	(0.232)	(0.232)	(0.233)
$Mock^{\text{STEM}}$		-1.915***	-1.774***			
		(0.189)	(0.199)			
$Mock^{\text{Non-STEM}}$					-2.217***	-2.200***
					(0.211)	(0.214)
Constant	-0.385*	1.434***	16.255***	-0.967***	0.593**	9.247***
	(0.205)	(0.300)	(3.042)	(0.209)	(0.283)	(3.204)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,196	5,196	5,196	5,196	5,196	5,196
R-squared	0.001	0.018	0.028	0.002	0.023	0.025
Panel B: Standardized score						
Female	0.025***	-0.034***	-0.038***	0.033***	0.045***	0.042***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
$Mock^{\text{STEM}}$		-0.215***	-0.210***			
		(0.007)	(0.007)			
$Mock^{\text{Non-STEM}}$					-0.064***	-0.063***
					(0.007)	(0.007)
Constant	-0.008	0.196***	0.680***	-0.023***	0.022**	0.313***
	(0.008)	(0.011)	(0.104)	(0.007)	(0.009)	(0.110)
Controls	No	No	Yes	No	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,196	5,196	5,196	5,196	5,196	5,196
R-squared	0.002	0.164	0.172	0.004	0.018	0.021

Note: OLS regressions. The sample consists of the top 50th percentile sample based on scores in the second mock exam in 2018-2019. The dependent variable is responses to competition. Columns 1-3 report results for STEM subjects and Columns 4-6 report results for non-STEM subjects. *Mock* refers to scores on the second mock exam. Scores are coded using percentile ranks in Panel A and using standardized scores in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.8: Pairwise correlation in response to competition across subjects

VARIABLES	(1) CO^{Math}	(2) $CO^{Physics}$	(3) $CO^{Chemistry}$	(4) $CO^{Chinese}$	(5) $CO^{English}$	(6) $CO^{History}$	(7) $CO^{Politics}$
CO^{Math}		0.338***	0.344***	0.178***	0.277***	0.199***	0.257***
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)
$CO^{Physics}$	0.353***		0.322***	0.158***	0.225***	0.205***	0.247***
	(0.009)		(0.009)	(0.009)	(0.009)	(0.011)	(0.011)
$CO^{Chemistry}$	0.389***	0.349***		0.184***	0.278***	0.229***	0.291***
	(0.009)	(0.009)		(0.009)	(0.010)	(0.010)	(0.011)
$CO^{Chinese}$	0.135***	0.115***	0.124***		0.117***	0.154***	0.199***
	(0.007)	(0.007)	(0.006)		(0.007)	(0.008)	(0.010)
$CO^{English}$	0.256***	0.199***	0.227***	0.143***		0.153***	0.207***
	(0.009)	(0.008)	(0.008)	(0.008)		(0.009)	(0.011)
$CO^{History}$	0.135***	0.134***	0.138***	0.138***	0.113***		0.227***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)		(0.009)
$CO^{Politics}$	0.109***	0.101***	0.109***	0.111***	0.095***	0.142***	
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	

Notes: Correlation coefficients between residualized response to competition in different subjects. The first three columns are STEM subjects and the last four columns are Non-STEM subjects. Scores are coded using percentile ranks. Robust standard errors are in parentheses.

Table D.9: Gender gaps in domain-general and comparative advantage in response to competition: by rural/urban middle school

	Overall		Above median		Below median	
	(1)	(2)	(3)	(4)	(5)	(6)
Urban students						
Panel A (1): PC_1						
Female	-0.097***	-0.107***	-0.069**	-0.082**	-0.134***	-0.137***
	(0.026)	(0.026)	(0.032)	(0.032)	(0.045)	(0.045)
R-squared	0.003	0.005	0.001	0.006	0.012	0.014
Panel A (2): $-PC_2$						
Female	-0.066***	-0.064***	-0.133***	-0.132***	0.034	0.036
	(0.018)	(0.018)	(0.023)	(0.023)	(0.032)	(0.032)
R-squared	0.001	0.001	0.005	0.005	0.002	0.003
Observations	11,982	11,982	7,451	7,451	4,531	4,531
Rural students						
Panel B (1): PC_1						
Female	-0.024	-0.034	0.004	0.001	-0.055	-0.061
	(0.033)	(0.033)	(0.061)	(0.061)	(0.039)	(0.039)
R-squared	0.005	0.011	0.015	0.021	0.006	0.010
Panel B (2): $-PC_2$						
Female	-0.150***	-0.149***	-0.197***	-0.196***	-0.128***	-0.130***
	(0.023)	(0.023)	(0.042)	(0.042)	(0.027)	(0.027)
R-squared	0.006	0.007	0.012	0.013	0.005	0.009
Observations	7,567	7,567	2,432	2,432	5,135	5,135
Controls	No	Yes	No	Yes	No	Yes

Note: OLS regressions. The dependent variable is the first principal component in Panel A(1) and B(1), and the negative of the second principal component in Panel A(2) and B(2). The sample consists of urban students in Panel A, and rural students in Panel B. Columns 1-2 report results for the full sample, Columns 3-4 report results for those with above median 2nd mock exam scores, and Columns 5-6 report results for those with below median 2nd mock exam scores. Scores are coded using percentile ranks. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.10: Gender gaps in domain-general and comparative advantage in response to competition: by family SES

	Overall		Above median		Below median	
	(1)	(2)	(3)	(4)	(5)	(6)
Non-poor students						
Panel A (1): PC_1						
female	-0.085***	-0.094***	-0.053*	-0.072**	-0.122***	-0.121***
	(0.022)	(0.022)	(0.031)	(0.031)	(0.033)	(0.033)
R-squared	0.001	0.005	0.002	0.006	0.002	0.008
Panel A (2): $-PC_2$						
Female	-0.093***	-0.095***	-0.145***	-0.152***	-0.040*	-0.044*
	(0.015)	(0.015)	(0.021)	(0.021)	(0.023)	(0.023)
R-squared	0.002	0.005	0.005	0.007	0.001	0.008
Observations	16,982	16,982	8,844	8,844	8,138	8,138
Poor students						
Panel B (1): PC_1						
Female	0.004	0.005	-0.020	-0.032	0.013	0.021
	(0.053)	(0.053)	(0.084)	(0.083)	(0.069)	(0.068)
R-squared	0.002	0.008	0.001	0.015	0.004	0.009
Panel B (2): $-PC_2$						
Female	-0.097**	-0.109***	-0.098	-0.105*	-0.086*	-0.098**
	(0.038)	(0.038)	(0.061)	(0.062)	(0.049)	(0.049)
R-squared	0.005	0.012	0.007	0.007	0.005	0.016
Observations	2,567	2,567	1,039	1,039	1,528	1,528
Controls	No	Yes	No	Yes	No	Yes

Note: OLS regressions. The dependent variable is the first principal component in Panel A(1) and B(1), and the negative of the second principal component in Panel A(2) and B(2). The sample consists of non-poor students in Panel A, and poor students in Panel B. Columns 1-2 report results for the full sample, Columns 3-4 report results for those with above median 2nd mock exam scores, and Columns 5-6 report results for those with below median 2nd mock exam scores. Scores are coded using percentile ranks. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.11: Determinants of STEM-track choice: Male students

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var = 1 if STEM track						
$STEM^{\text{Mock}}$	0.931***	0.913***	0.633***	0.623***	0.581***	0.581***
	(0.043)	(0.044)	(0.046)	(0.062)	(0.064)	(0.064)
CA_{Mock}		0.127*	0.106	0.109*	0.140*	0.143*
		(0.052)	(0.054)	(0.056)	(0.056)	(0.056)
$High\_Perf\_Females$			-0.004***	-0.004***	-0.004***	-0.004***
			(0.000)	(0.000)	(0.000)	(0.000)
RC^{STEM}					0.871***	0.880***
					(0.109)	(0.109)
$RC^{non-STEM}$						-0.042
						(0.093)
Controls	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.223	0.225	0.214	0.220	0.253	0.253
Observations	3,219	3,219	2,548	2,548	2,548	2,548

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of male students. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.12: Determinants of STEM-track choice: Female students

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var = 1 if STEM track						
$STEM^{\text{Mock}}$	1.424***	1.289***	1.088***	1.422***	1.371***	1.395***
	(0.056)	(0.059)	(0.081)	(0.129)	(0.156)	(0.148)
CA_{Mock}		0.558***	0.797***	0.661***	0.897***	0.961***
		(0.091)	(0.115)	(0.122)	(0.126)	(0.128)
$High\_Perf\_Females$			-0.006***	-0.007***	-0.007***	-0.007***
			(0.001)	(0.001)	(0.001)	(0.001)
RC^{STEM}					1.903***	2.028***
					(0.211)	(0.204)
$RC^{non-STEM}$						-0.637***
						(0.190)
Controls	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.173	0.181	0.191	0.198	0.234	0.238
Observations	3,233	3,233	2,523	2,523	2,523	2,523

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of female students. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.13: Determinants of STEM-track choice: elite high school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var = 1 if STEM track							
$Female$	-0.228***	-0.183***	-0.144***	-0.138***	-0.142***	-0.142***	-0.106***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
$STEM^{\text{Mock}}$		1.036***	0.869***	0.684***	0.726***	0.683***	0.702***
		(0.080)	(0.078)	(0.082)	(0.088)	(0.090)	(0.089)
CA_{Mock}			0.607***	0.752***	0.733***	0.750***	0.814***
			(0.087)	(0.090)	(0.091)	(0.092)	(0.090)
$High\_Perf\_Females$				0.001*	0.001*	0.001*	0.001*
				(0.000)	(0.000)	(0.000)	(0.000)
PC^1						0.013	0.019*
						(0.009)	(0.009)
$-PC^2$							-0.090***
							(0.009)
Controls	No	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.080	0.195	0.214	0.202	0.204	0.205	0.237
Observations	3,085	3,085	3,085	3,042	3,042	3,042	3,042

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of students from elite high schools. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.14: Determinants of STEM-track choice: non-elite high school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var = 1 if STEM track							
$Female$	-0.253***	-0.191***	-0.176***	-0.192***	-0.200***	-0.195***	-0.160***
	(0.017)	(0.019)	(0.020)	(0.028)	(0.028)	(0.029)	(0.030)
$STEM^{\text{Mock}}$		1.544***	1.494***	1.464***	1.487***	1.377***	1.467***
		(0.070)	(0.075)	(0.118)	(0.121)	(0.115)	(0.120)
CA_{Mock}			0.137	-0.075	-0.085	0.022	-0.010
			(0.070)	(0.101)	(0.101)	(0.103)	(0.105)
$High\_Perf\_Females$				-0.017***	-0.017***	-0.018***	-0.018***
				(0.001)	(0.001)	(0.001)	(0.001)
PC^1						0.119***	0.130***
						(0.012)	(0.013)
$-PC^2$							-0.091***
							(0.014)
Controls	No	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.048	0.205	0.206	0.379	0.382	0.419	0.434
Observations	3,367	3,367	3,367	2,029	2,029	0 2,029	2,029

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of students from non-elite high schools. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.15: Determinants of STEM-track choice: Male students

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var = 1 if STEM track						
$STEM^{\text{Mock}}$	0.931***	0.913***	0.633***	0.623***	0.544***	0.574***
	(0.043)	(0.044)	(0.046)	(0.062)	(0.062)	(0.063)
CA_{Mock}		0.127*	0.106	0.109*	0.140*	0.119*
		(0.052)	(0.054)	(0.056)	(0.056)	(0.055)
$High\_Perf\_Females$			-0.004***	-0.003***	-0.004***	-0.004***
			(0.000)	(0.000)	(0.000)	(0.000)
PC^1					0.033***	0.037***
					(0.007)	(0.007)
$-PC^2$						-0.051***
						(0.007)
Controls	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.223	0.225	0.214	0.220	0.232	0.253
Observations	3,219	3,219	2,548	2,548	2,548	2,548

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of male students. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.16: Determinants of STEM-track choice: Female students

	(1)	(2)	(3)	(4)	(5)	(6)
Dep var = 1 if STEM track						
$STEM^{\text{Mock}}$	1.424***	1.289***	1.088***	1.422***	1.259***	1.339***
	(0.056)	(0.059)	(0.081)	(0.129)	(0.139)	(0.146)
CA_{Mock}		0.558***	0.797***	0.661***	0.774***	0.864***
		(0.091)	(0.115)	(0.122)	(0.123)	(0.124)
$High\_Perf\_Females$			-0.006***	-0.007***	-0.007***	-0.007***
			(0.001)	(0.001)	(0.001)	(0.001)
PC^1					0.067***	0.076***
					(0.013)	(0.014)
$-PC^2$						-0.124***
						(0.012)
Controls	No	No	No	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.173	0.181	0.191	0.198	0.209	0.234
Observations	3,233	3,233	2,523	2,523	2,523	2,523

Note: Probit regressions, marginal effects reported. The dependent variable equals 1 if the STEM track is chosen; 0 if non-STEM track is chosen. The sample consists of female students. The sample consists of the 2015, 2016, and 2018 cohorts. The 2017 cohort has been excluded due to the unavailability of HSEE data, as indicated in Appendix Table B.1. $STEM^{Mock}$ is performance in the second mock HSEE in STEM. CA^{Mock} is the comparative advantage in STEM versus non-STEM subjects in the 2nd mock HSEE. All scores are coded as percentile ranks. *High_Perf_Females* is the proportion female in the top 20% of each high school class. RC^{STEM} is residualized response to competition in STEM in the HSEE. $RC^{non-STEM}$ is residualized response to competition in non-STEM in the HSEE. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.17: Correlation between response to competition in HSEE and the CEE (by gender)

VARIABLES	(1) $RC^{Math-CEE}$	(2)	(3) $RC^{Chinese-CEE}$	(4)	(5) $RC^{English-CEE}$	(6)
Panel A: Female = 1						
$RC^{Math-HSEE}$	0.292***	0.285***				
	(0.038)	(0.038)				
$RC^{Chinese-HSEE}$			0.276***	0.263***		
			(0.035)	(0.035)		
$RC^{English-HSEE}$					0.123***	0.116***
					(0.034)	(0.034)
Constant	-0.343	19.924**	0.493	64.766***	0.375	12.096*
	(0.673)	(8.678)	(0.825)	(10.943)	(0.490)	(6.259)
Controls	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,005	2,005	2,005	2,005	2,005	2,005
R-squared	0.033	0.036	0.028	0.050	0.007	0.012
Panel B: Female = 0						
$RC^{Math-HSEE}$	0.242***	0.241***				
	(0.046)	(0.046)				
$RC^{Chinese-HSEE}$			0.311***	0.309***		
			(0.040)	(0.040)		
$RC^{English-HSEE}$					0.108***	0.108***
					(0.030)	(0.030)
Constant	3.307***	10.504	-1.313	43.893***	-0.694	9.288
	(0.930)	(10.094)	(1.279)	(12.841)	(0.770)	(7.365)
Controls	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,755	1,755	1,755	1,755	1,755	1,755
R-squared	0.019	0.022	0.035	0.042	0.008	0.011

Note: Correlation coefficients between residualized response to competition in the HSEE and in the CEE. Scores are coded using percentile ranks. The sample consists of female students in Panel A and male students in Panel B. Controls include age at HSEE, a dummy for graduating from an urban middle school, and a dummy for being from a poor family. Robust standard errors are in parentheses.

Table D.18: Response to competition (CEE) - excluding controls for mock exam scores

VARIABLES	(1) Math	(2) Chinese	(3) English
Female	-0.012*	-0.023**	-0.006
	(0.006)	(0.009)	(0.004)
Constant	0.007	0.009	0.025***
	(0.008)	(0.010)	(0.005)
Cohort FE	Yes	Yes	Yes
Observations	3,764	3,764	3,764
R-squared	0.002	0.002	0.003

Note: OLS regressions. The dependent variable is response to competition in the CEE. Scores are coded using percentile ranks. Robust standard errors are in parentheses.

Table D.19: Response to competition (CEE): by track

VARIABLES	(1) Math	(2) STEM Track Chinese	(3) English	(4) Math	(5) Non-STEM track Chinese	(6) English
Female	-0.024***	0.039***	0.014***	-0.032***	0.037***	0.024***
	(0.007)	(0.009)	(0.005)	(0.010)	(0.013)	(0.009)
$Math^{Mock}$	-0.295***			-0.327***		
	(0.015)			(0.024)		
$Chinese^{Mock}$		-0.639***			-0.539***	
		(0.018)			(0.027)	
$English^{Mock}$			-0.160***			-0.203***
			(0.011)			(0.021)
Constant	0.344***	0.713***	0.217***	0.278***	0.818***	-0.022
	(0.084)	(0.102)	(0.059)	(0.101)	(0.121)	(0.076)
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,609	2,609	2,609	1,155	1,155	1,155
R-squared	0.137	0.340	0.076	0.178	0.288	0.123

Note: OLS regressions. The dependent variable is response to competition in the CEE. Columns 1-3 report results for STEM track students, Columns 4-6 report results for Non-STEM track students. *Mock* refers to scores on the 4th and last mock CEE. Scores are coded using percentile ranks. Controls include age at HSEE, a dummy for graduating from an elite high school, and a dummy for being from a poor family. Robust standard errors are in parentheses.