Gender Composition and High-Stakes Cognitive Performance: Evidence from a Quasi-Randomized Experiment

Xuan Li
Boston University

Xiang Zhou
Xiangtan University

December 6, 2024

Motivation

- Debates surrounding diversity (e.g., gender composition) in schools remain contentious.
- Concerns about diversity often highlight the challenges that intensive daily interactions pose to underrepresented or discriminated-against groups.
- This paper: Whether the mere presence of other groups can affect educational outcomes?

Setting: National College Entrance Examination (*Gaokao*)

- Organization of the *Gaokao*:
  - → The exam spans two days, and students complete the tests individually.
  - → Students are randomly assigned to test rooms.
- A high-stakes exam
  - → College admission is determined solely by the score achieved in the *Gaokao*
  - → Attending an elite college offers substantial benefits.
- **Research Question:** Does the proportion of males in the test room affect the performance of female students during a high-stakes, non-interactive exam?

Main Findings

- An increased presence of male students in the test room decreases the performance of female students but has no effect on male students.
- Channel: Stereotype threat.
  - → The presence of males is a common cue for triggering stereotype threat in lab experiments.
  - → The effect of gender composition varies based on individuals' beliefs about stereotypes, particularly among those who:
    - have low performance in STEM subjects,
    - have high performance in non-STEM subjects, and
    - come from high school classes with a higher proportion of male students.

Related Literature

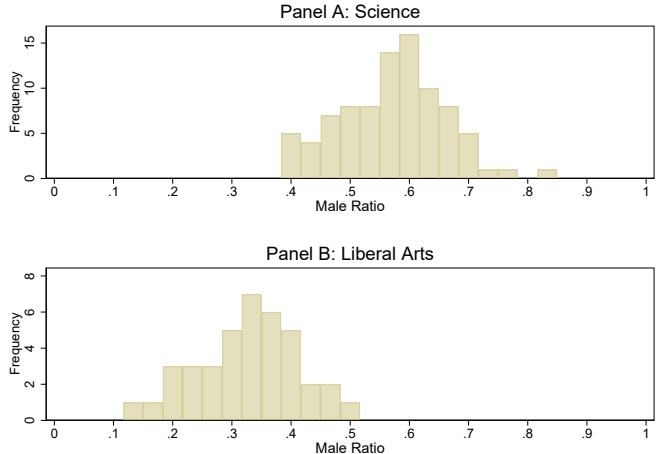
- Diversity and Educational Outcomes
  - → The first study to demonstrate **the passive effects of gender composition** on academic outcomes.

- Stereotype Threat
  - → The first study conducted in **a real-world, high-stakes setting with a large sample size**.
  - → We demonstrate that vulnerability to stereotype threat is influenced by individuals' beliefs about stereotypes, which helps to reconcile inconsistencies in the literature.

Administrative Student-level Data

- A county in the Central China
- Sample period: 2019
- 127 test rooms, with 30 students in each room.
- Two academic tracks:
  - → Science
    - 1509 male students and 1131 female students
  - → Liberal arts
    - 382 male students and 788 female students.

127 Test Rooms: 88 Science, 39 Liberal Arts



The histogram of the test room's gender composition aligns with the expected pattern of a binomial distribution. ▸ Histogram of χ^2 Statistic

Administrative Student-level Data

- Variables
  - → Total *Gaokao* score (only available for these obtain college admission)
  - → Mock exam score (by subjects)
  - → High school, class
  - → Test center, test room
  - → Individual characteristics: gender, ethnicity, date of birth, student leader, etc.

Balance

VARIABLES	(1) Han	(2) Age	(3) CYL	(4) Leader	(5) Urban	(6) Mock Exam Score	(7) Admission
Panel A: Male							
MaleRatio	-0.00302	0.264	-0.0194	0.0153	-0.144	-0.147	0.0562
	(0.0185)	(0.198)	(0.0657)	(0.0881)	(0.100)	(0.289)	(0.119)
Track FEs	X	X	X	X	X	X	X
Panel B: Female							
MaleRatio	0.00397*	-0.252	-0.00602	0.0439	0.00366	-0.0644	-0.000647
	(0.00235)	(0.202)	(0.0390)	(0.0812)	(0.0938)	(0.235)	(0.104)
Track FEs	X	X	X	X	X	X	X

Preferred Specification

$$StdScore_i = \beta_1 MaleRatio_r + \gamma Ind_i + \phi TR_r + \pi_h + \pi_c + \epsilon_i$$

- $StdScore_i$: the standardized total score of student i in the *Gaokao*.
- $MaleRatio_r$: the proportion of male students in the test room r.
- Ind_i: individual-level characteristics
- TR_r: test room characteristics
- π_h: **high school class fixed effects**
- π_c: test-center fixed effects
- Standard errors are clustered at the test-room level.

Results

	Panel A: Male			Panel B: Female			Panel C: All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES		Std Score			Std Score			Std Score
MaleRatio (β_1)	0.0918	0.0931	0.0609	-0.350*	-0.607**	-0.396**	0.105	0.220
	(0.198)	(0.574)	(0.178)	(0.185)	(0.281)	(0.172)	(0.185)	(0.444)
Female (β_2)							0.241	0.672***
							(0.148)	(0.145)
MaleRatio*Female (β_3)							-0.517**	-0.660
							(0.248)	(0.464)
$H_0 : \beta_1 + \beta_3 = 0$							-0.412**	-0.441*
							(0.191)	(0.260)
Observations	1,095	258	1,354	871	614	1,485	1,966	874
Ind Controls	X	X	X	X	X	X	X	X
Test-Room Controls	X	X	X	X	X	X	X	X
Test-Center FEs	X	X	X	X	X	X	X	X
Class FEs	X	X	X	X	X	X	X	X
Sample	Science	Liberal Arts	All	Science	Liberal Arts	All	Science	Liberal Arts

Economic Significance

- Science track:
  - → a test room with 15 male students → 20 male students
  - → ↓ 7.3 points (0.167 * -0.412 * 105.8)
  - → Ranking (at most) ↓ 6,400 out of 210,000 students

- Liberal arts track:
  - → a test room with 10 male students → 15 male students
  - → ↓ 6.8 points (0.167 * -0.441 * 92.3)
  - → Ranking (at most) ↓ 4,100 out of 150,000 students

Stereotype Threat

- The fear or anxiety of confirming a negative stereotype about one's social group
- An prevalent stereotype is the belief that women are bad at STEM subjects.
  - → 60% of male students and 50% of female students in China agree with this.
- The presence of males is a common cue for triggering stereotype threat in lab experiments (Inzlicht and Ben-Zeev, 2000)

Who Holds Stereotypes?

- China Education Panel Survey (2013)
  - → Do you agree that boys are better at mathematics than girls?
- Sample: Grade 9 female students
- Individual-level features
  - → STEM exam score
  - → Non-STEM exam score
- Class-level features
  - → Class male ratio
  - → Whether males perform better than females in STEM subjects

VARIABLES	(1)	(2)	(3)
	Gender Stereotype		
STEM Exam std score	-0.148***	-0.163***	-0.167***
	(0.0147)	(0.0154)	(0.0145)
Non-STEM Exam std score	0.0690***	0.0783***	0.0859***
	(0.0168)	(0.0168)	(0.0166)
Class Male Ratio (Above Median)			0.0378*
			(0.0213)
Class: Males Performing Better in STEM			0.0332
			(0.0370)
Observations	3,548	3,547	3,548
R-squared	0.045	0.151	0.110
Class FEs	.	X	.
School FEs	.	.	X

VARIABLES	(1) Std Score	(2) Std Score	(3) Std Score	(4) Std Score
MaleRatio	-0.418***	0.000745	-0.140	0.252
	(0.160)	(0.143)	(0.223)	(0.594)
STEM Mock Exam Std Score	0.363***			
	(0.0796)			
Non-STEM Mock Exam Std Score		1.044***		
		(0.0652)		
MaleRatio*STEM Mock Exam Std Score	0.304*			
	(0.162)			
MaleRatio*Non-STEM Mock Exam Std Score		-0.631***		
		(0.132)		
MaleRatio*HS Class Male Ratio (Above Median)			-0.728**	
			(0.333)	
MaleRatio*HS Class: Males Performing Better in STEM				-1.445
				(1.144)
Observations	1,423	1,423	1,485	1,458
R-squared	0.764	0.812	0.645	0.634
Ind Controls	X	X	X	X
Test-Room Controls	X	X	X	X
Test-Center FEs	X	X	X	X
Class FEs	X	X	X	X
Sample	All	All	All	All

Alternative Interpretations

- Gender Differences in Competition
  - → Levels of competitiveness are the same
    - Identical exams
    - Same pool of competitors
  - → Inconsistent with heterogeneity analysis of gender stereotype endorsement

Conclusions

- **The effect of gender composition goes beyond social interaction and active peer influence.**
- **Policy relevance:**
  - → Standardized tests are widely used and carry high stakes.
  - → Most testing environments are mixed-gender.
  - → Gender stereotypes are prevalent worldwide.
- **Policy recommendation: Implement a single-gender test environment.**

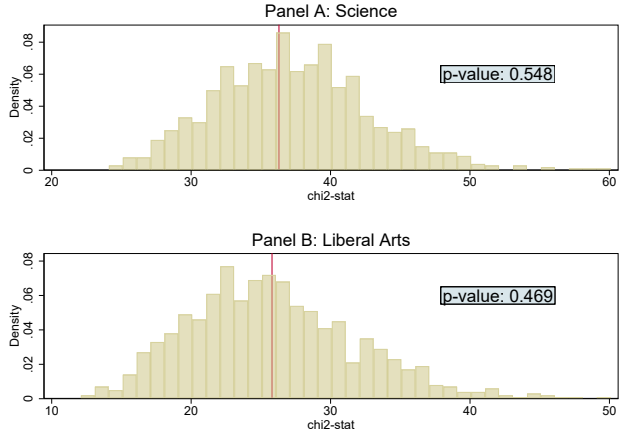
Thank You!

Xuan Li

Job Market Candidate, Boston University

Email: xuanli@bu.edu Website: www.xuan-li.net

Histogram of χ^2 Statistic

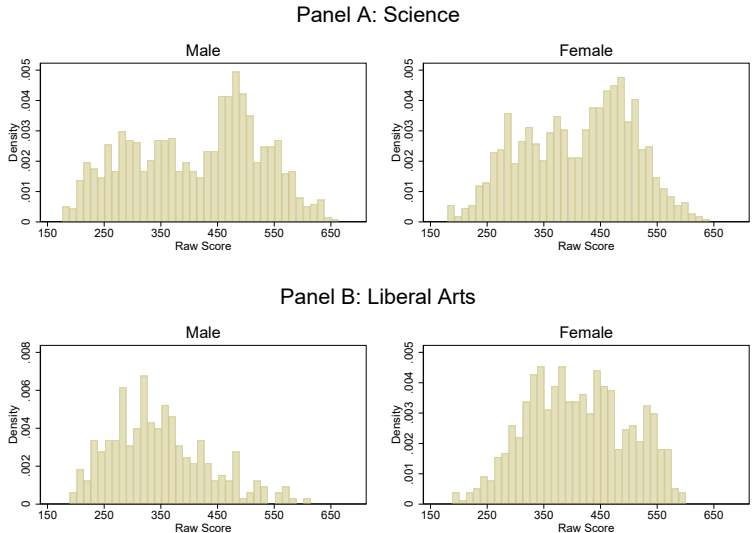


Go Back

Gender Composition and Other Test Room Characteristics

VARIABLES	(1) HanRatio	(2) Average Age	(3) CYLRatio	(4) ClassLeaderRatio	(5) UrbanRatio
MaleRatio	-0.00181	0.114	-0.0385	0.0184	-0.0501
	(0.00919)	(0.141)	(0.0381)	(0.0522)	(0.0555)
Observations	127	127	127	127	127
Track FEs	X	X	X	X	X

Raw Score


Panel A: Science
Male
Female
Density
Raw Score
0
.001
.002
.003
.004
.005
150
250
350
450
550
650
Panel B: Liberal Arts
Male
Female
Density
Raw Score
0
.002
.004
.006
.008
.001
.002
.003
.004
.005
150
250
350
450
550
650