Adding Up Peer Beliefs:

Experimental and Field Evidence on the Effects of Peer Influence on Math Performance

Abstract

We study how gendered beliefs about intellectual abilities transmit through peers and differentially impact girls' academic performance relative to boys. Study 1 (N = 8029; 208 classrooms) exploits randomly assigned variation in the proportion of a child's middle school classmates who believe boys are innately better than girls at learning math. An increase in exposure to peers who report this belief generates losses for girls and gains for boys regarding math performance. This peer exposure also increases children's likelihood of believing the gender-math stereotype, increases perceived difficulty of math and reduces aspirations among girls. Study 2 (N = 547) provides proof-of-concept that activating a gender-math performance gap among college students reduces women's math performance, but not verbal performance. Men's task performance is not affected. Our findings highlight how the prevalence of stereotypical beliefs in one's ambient and peer environment, even when readily contradictable, can shape children's beliefs and academic ability.

Word count = 150

Keywords: peer influence, gender-math stereotype, gender gap, belief formation, STEM aspiration

Significance Statement

Gender disparities in STEM occupations and interest persist in many societies and coincide with the stereotypical belief that men are inherently better at learning mathematics than women. This gender-math stereotypical belief may be transmitted through peers, affecting one's belief formation and aspirations, and further affecting one's math abilities. From a large-scale natural experiment on middle-school children and a pre-registered laboratory experiment, we found gendered beliefs in intellectual abilities had both immediate and prolonged effects on female students' academic performance and belief formation. These effects were potentially transmitted through one's peer groups, where same-gender peers had a more pronounced impact than different-gender peers. These findings highlight the importance of one's ambient and peer environment on the propagation of gendered beliefs and their cumulative impact on intellectual abilities and psychological outcomes.

Word count = 129

Adding Up Peer Beliefs:

Experimental and Field Evidence on the Effect of Peer Influence on Math Performance

A persistent gender gap exists in the science, technology, engineering, and math (STEM) workforce around the world, despite the progress in gender equality in overall educational attainment (Riegle-Crumb et al., 2011; OECD, 2015). Women remain substantially under-represented in the most in-demand and high-paying STEM domains: Only 28% of employed scientists and engineers are women (National Academy of Sciences, 2007; National Survey of College Graduates, 2015). Coinciding with the under-representation of women is the prevailing notion that men are inherently better than women at learning mathematics. This stereotypical belief persists in many countries despite the fact that women often perform as well as or better than men in K-12 math assessments (Eble & Hu, 2022; Gong et al., 2018; Jayachandran, 2015). The gender gap in STEM has clear unfavorable consequences: First, it is highly relevant to gender wage inequality in the workforce as STEM jobs tend to be more lucrative; second, it reflects the under-utilization of talents in today's increasingly high demand for STEM workers (Liu, 2018; Perry et al., 2012). Therefore, it is imperative to understand the forces that perpetuate gender inequality in STEM fields, including the development roots of the gender-math stereotypical belief and how it transmits among children and affects their belief formation and aspiration.

The present paper demonstrates how the early emerging stereotypical belief that "boys are innately better at learning math than girls" may transmit in children's peer environment and change their math ability and psychological outcomes. We utilized a natural experiment among middle-school children who were randomly assigned to different classrooms. We show that the prevalence of gendered belief in math ability in a child's peer environment influences their subsequent performance in midterm math exams as well as non-cognitive outcomes. Next, we conducted a pre-registered laboratory experiment to provide proof-of-concept that the saliency of a gender-math performance gap can pose an immediate effect on women's math performance. The combination of the field and lab data allows us to provide rigorous evidence that the level of gender-math stereotype in the ambient environment can have both immediate and cumulative effects on students' math achievement.

We focus on the influence of peer beliefs in children's and young adults' ambient environment because the roots of gender inequality in STEM aspiration and attainment stem further back. Research shows that gender stereotypes about intelligence emerge early and affect children's interests (Bian et al., 2017). Common stereotypes associate math and high-level intellectual ability with men more than women, and stereotypical beliefs such as "boys are innately better in math than girls" may discourage girls from pursuing math and other STEM-related domains. In particular, educational and occupational aspirations begin to crystallize in early adolescence, coinciding with the development of increasingly salient gender identity and gender roles (Bandura et al., 2001; Eccles & Roeser, 2011). Young people are subject to a multitude of messages from the ambient environment regarding what is appropriate and expected from their own gender group. It is also during adolescence when young people begin to move away from parental influence and become increasingly susceptible to peer influence (Reigle-Crumb & Morton, 2017; Wentzel, 2012). The peer influence might be even more pronounced for girls than boys as girls are socialized to be more aware of and sensitive to others' opinions (Gilligan, 1982; Beutel & Marini, 1995).

Research on the stereotype threat investigates how gender stereotypes and bias can function within local environments to deter women's STEM interest and achievement (Steele, 1991; 1997; 2010; Shih et al., 1999). It is hypothesized that the gender-math gap persists in part because widely known negative stereotypes about women's math abilities put female students in situations that put pressure on them about how they would be evaluated by others (Lewis & Michalak, 2019). These situations are thought to be threatening enough as to hamper academic performance (Steele & Aronson, 1995). Given the prominence of stereotype threat, or social identity threat more generally, in attributing to the gender disparity in STEM, a large number of replication studies have been conducted to refine the theory and calibrate the magnitude of the effects. According to meta-analyses (Flore & Wicherts, 2015; Nguyen & Ryan, 2008), large and significant findings were concentrated among earlier and small-sample studies, whereas recent reanalysis and replication attempts failed to replicate earlier findings (e.g., Finnigan & Corker, 2016; Ganley et al., 2013; Stoet & Geary, 2012; Zigerell, 2017; Shewach et al., 2019). Notably, many replication attempts do not cast doubt on the existence of the effect per se, but provided clarification on the robustness of the phenomenon in different contexts. In addition, most studies focus on the experience of the individuals targeted with negative stereotypes. Less is known about how these stereotypes may transmit through social environment and group interactions and subsequently impact academic performance.

In this research, we focus on the transmission of stereotypical beliefs in children's peer environment and tests the impact of gender-math stereotype on students' academic and psychological outcomes in a context where the widely held belief is readily contradictable. We predict a causal relationship between the prevalence of gender-math stereotypes in one's peer environment and girls' effort and math performance. The theoretical intuition is that the stereotypical belief—"boys are better in math than girls"—can transmit in one's ambient environment through peer interactions. Girls must manage sporadic comments from peers about the innately inferior math learning ability of their gender group, make sense of what is expected of them to study, and juggle doubts about their own capabilities. At the moment, the saliency of negative messages can be threatening and preoccupations with these negative stereotypes can be distracting. Over time, girls in a high stereotype environment can be less motivated and less able to exert effort in math study and subsequently fare worse in math tests.

We tested these theoretical intuitions using two different but complementary designs. Studying peer influence of gender-math stereotype is challenging because it is ethically unjustifiable to randomly expose some children to a greater level of stereotype and bias for the purpose of studying them because of the documented harm to the disadvantaged group. Our first study sidesteps this concern and exploits the random assignment of children to the classroom to generate quasi-experimental evidence of peer influence on children's academic performance. We focus on a mechanism that is under-studied: the peer influence of gender-math stereotypes over time. Because the natural experiment does not offer direct momentary causal evidence of a stereotypical belief, we bolstered the causal mechanism by testing the immediate effect of exposing to gender-math stereotypes in a laboratory. As such, the two studies in this paper are highly complementary. The natural experiment demonstrates the phenomenon in the real world and provides us an estimate of the long-term effect of classroom-level gender-math stereotypes on girls' versus boys' academic performance, while the laboratory study provides a proof-of-concept that activating gender-math stereotypes can cause immediate impact on female students' math performance.

The Natural Experiment

Method

The first study utilized a sample of schools in the China Education Panel Survey (CEPS), a large-scale and nationally representative sample of middle school students in mainland China.[1] The CEPS uses a stratified, multistage sample design covering 19,487 students from 438 middle schools in 112 counties (districts). In each sampled county

[1] The CEPS is the first and largest nationally representative longitudinal survey to focus on secondary school students and teachers in mainland China.

(district), four middle schools with Grades 7 and 9 were chosen, and four classes were surveyed from each middle school, including two Grade 7 classes and two Grade 9 classes. To identify the causal impact, we limited the sample to schools where students are randomly assigned to classrooms. Our sample contains 67 middle schools, 208 classrooms, and 8,029 students covering 26 counties (49.3% female; average age is 13 years; 10.7% ethnic minorities). In the supplementary materials, we present additional statistical evidence to show that schools in our sample do assign students randomly, as reflected by predetermined student performance and academic performance characteristics.

In this natural experiment, we did not experimentally induce the salience of gender gap in math performance, but we used the quasi-experimental variation in actual classroom-level belief that boys are innately better than girls at learning math. This research design allows for the study of the impact of being assigned to a peer environment (i.e., a class) with different peer characteristics (Eble & Hu, 2022; Feld & Zölitz, 2017; Sacerdote, 2011). At the beginning of middle school in China, students are assigned to classes and remain in the same class for the next three years. In the same class, students interact extensively, both academically and outside of class. Every lecture is delivered according to the same schedule as the class. Additionally, students participate in self-study sessions as well as outdoor activities together. We predict that the classroom-level *Gender-Math stereotype* will affect various academic and nonacademic social interactions and subsequently impact girls' math performance in an accumulative way.

In the CEPS sample, middle school students in the 7th and 9th grades completed a questionnaire about their beliefs and aspirations at the beginning of their academic year, before any formal tests took place. Specifically, we obtained students' gender-math beliefs from their responses to the following question, "Do you agree that boys' natural ability in studying math is greater than that of girls?" The wording of this question refers to the innate math ability of each gender, not just the relative performance of boys and girls in the child's current school or classroom. This question enables us to generate a classroom-level variable summarizing the beliefs of the child's peers in their randomly assigned classroom. We quantify the *Gender-Math stereotype*, the key independent variable, as the proportion of students' peers in their classroom who believe that boys' innate math ability (Bm) is superior to girls' innate math ability (Gm). There were a lot of naturally occurring variations in the aggregate level of stereotypical belief: in some classrooms, only 13.33% in others up to 91.89% (Appendix Figure 1 shows the peer Gender-Math stereotype between classes). These variations are not related to gender compositions of a classroom.

All middle schools in our sample administer midterm exams in the middle of the fall semester. The CEPS student data include administrative data on the child's test scores in three core subjects: Chinese, English, and Mathematics. All students within a grade of a school take the same midterm exam in each subject, which is graded centrally at the school level. Accordingly, student test scores across classes within a particular grade are comparable at each school. In addition to administrative data on academic performance, we also use measures on students' extra-curricular participation and students' self-reports of confidence, aspirations, and perception of teachers' behaviors.

Analysis Strategy. In our empirical analysis, we focus on estimating the effect of the proportion of peers who believe that boys are innately better at learning math than girls on a child's academic and psychological outcomes, and how the effect varies with the child's gender. Our identification strategy exploits the random variation between classrooms in a given grade within a school in the proportion of peers who hold this gender-math stereotypical belief. The random classroom assignment allows us to control for other external variables that might affect the children's academic performance. To examine the impact of peer gender-math beliefs in the classroom on students' academic outcomes, we used the fixed-effect linear model as follows:

$$Y_{icgs} = \beta_0 + \beta_1 PB_{icgs} * Female_{icgs} + \beta_2 PB_{icgs} + \beta_3 Female_{icgs} + \phi X_{icgs} + \tau W_{cgs} + \lambda_s + \lambda_g + \varepsilon_{icgs}$$

Where Y_{icgs} is the outcome variable for student *i* in class *c* in grade g in school s. PB_{icgs} is the *Gender-Math stereotype* in a child's classroom, defined as the proportion of student *i*'s peers in their classroom who believe that boys have innately greater math ability than girls—standardized to be mean 0 and standard deviation 1. $Female_{icgs}$ indicates whether or not the student i is female. To increase the accuracy of our estimations, we also control for student characteristics X_{icgs} and teacher characteristics W_{cgs} that are not affected by class assignment. λ_s is the School fixed effect, λ_g is the Grade fixed effect and ε_{icgs} is the error term. We cluster the standard errors at the class level, accounting for correlation in outcomes among students in the same class. See the supplementary materials for detailed demographic descriptions and balance test.

The coefficient we are most interested in is β_1, which measures how a one standard deviation (*SD*) increase in the proportion of peers who believe boys have innately superior math ability than girls impacts girls' outcomes versus boys'. In other words, it represents the impact of peer beliefs in the class on the "gender gap" (c.f., Muralidharan & Sheth, 2016). The coefficient β_2 measure how boys' outcomes are affected by a one *SD* increase in peers who believe that boys have innately superior math skills. Because the classrooms are randomly assigned, β_1 and β_2 represent unbiased estimates of the peer gender-math stereotypes effect (see the supplementary materials for robustness checks for causal identification).

The Natural Experiment Results

Academic Performance. We first examined the effect of peer gender-math stereotypes on test scores of three core subjects—Chinese, mathematics, and English (see Table 1 for main results and Appendix Table A.4 for results on simple effects). In this nationally representative sample, girls outperform boys on average in all three core subjects, consistent with prior empirical findings in China (Akabayashi et al., 2020; Gu & Yeung, 2021). However, even in this setting where girls' demonstrated test scores were significantly higher than boys, a considerable proportion of students still believed that "boys are innately better in math than girls" (*M* = 53.3%, *SD* = 0.499; see Appendix Table A.1).

It is estimated that girls' test scores in math, relative to boys', worsened by 0.894 standard deviations upon being exposed to peers who believed that boys had innately superior abilities in math. On the other hand, a one standard deviation increase in the stereotypical peer belief measure led to an increase of 0.439 standard deviations in the boy's standardized math scores. Interestingly, the coefficients of $PB_{icgs} * Female_{icgs}$, indicating the effect of peers' stereotypical gender-math beliefs, were only statistically significant for math but insignificant for Chinese and English. The results indicated that girls' math performance would decline as more of their peers believed boys had innately superior abilities in math, and the gender gap would widen. However, we did not observe this gender difference in Chinese or English test scores (see Columns 1 to 3 in Table 1).

We next explore how these patterns are differentially impacted by peers of different genders. Specifically, we calculated two class-specific measures of the proportion of peers who believe that boys have innately superior abilities in math, one for girl peers and one for boy peers. As children with the same gender identity are more likely to interact in this age group, peers of the same gender should be more influential in transmitting beliefs than peers of different gender (Currarini et al., 2009; Eble & Hu, 2022). According to Column 4 and 5 in Table 1, same-gendered peers' beliefs had a smaller impact on math performance than beliefs of peers from a different gender group. Table A.5 in the supplementary material shows these estimates using raw test scores as outcomes.

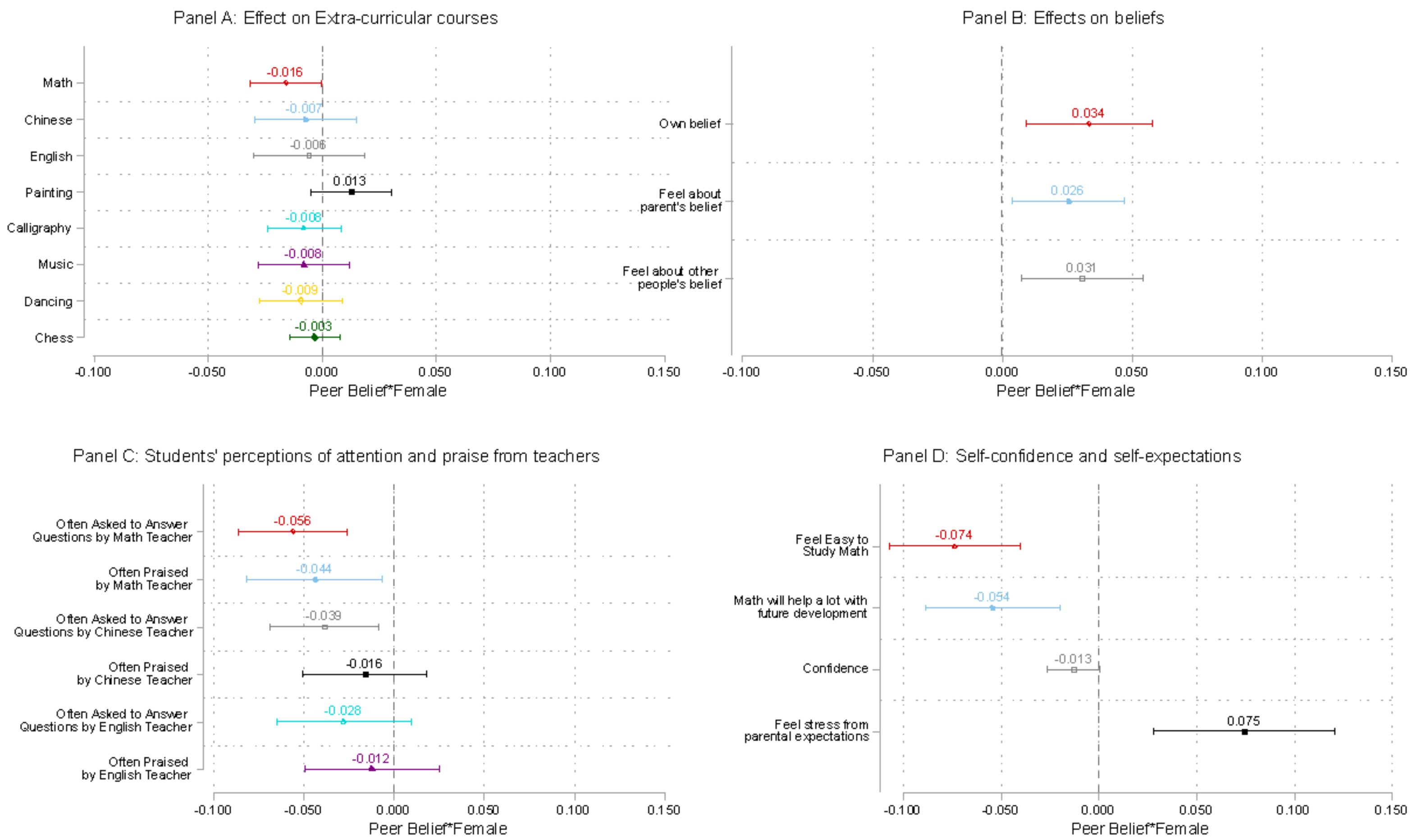
Extra-Curricular Course Enrollment. Next, we examine whether children's participation in extra-curricular activities is affected by the proportion of peers who hold the gender-math stereotypical belief. We find that having a greater proportion of peers who hold the gender-math stereotypical belief only lowered girls' participation in extra-curricular mathematics courses (b = -0.016, SE = 0.010, $p < .1$). However, we did not observe this type of gender disparity in participation in other extra-curricular courses, such as Chinese, English, painting, etc. (see Figure 1 and Appendix Table A.6).

Table 1. *Effect on Academic Performance.*

	(1) Math	(2) Chinese	(3) English	(4) Math	(5) Math
Peer Belief*Female	-0.894***	-0.334	-0.392		
	[0.259]	[0.253]	[0.253]		
Own-gendered Peer Belief*Female				-0.656**	
				[0.261]	
Other-gendered Peer Belief*Female					-1.133***
					[0.260]
Peer Belief	0.439***	0.364*	-0.091	0.211	0.649***
	[0.140]	[0.214]	[0.196]	[0.189]	[0.188]
Female	1.216***	5.687***	5.301***	1.148***	1.167***
	[0.267]	[0.240]	[0.241]	[0.296]	[0.291]
School fixed effects	YES	YES	YES	YES	YES
Grade fixed effects	YES	YES	YES	YES	YES
Baseline Controls	YES	YES	YES	YES	YES
Observations	8,029	8,030	8,029	8,029	8,029
Adjusted R-squared	0.074	0.123	0.124	0.073	0.074

Note: Each column represents the coefficient from a separate regression. For columns (1), (4) and (5), the dependent variable is the standardized math score. For column (2), the dependent variable is the standardized Chinese score. For column (3), the dependent variable is the standardized English score. Standard errors are clustered at the class level. SEs are in parentheses. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Figure 1. *Effect of Peer Influence on Extra-Curricular Participation and Psychological Outcomes.*



Note: This figure provides the coefficient values and 90% confidence intervals for the effect of peer influence on extra-curricular participation and belief outcomes. Each dot represents the coefficient from a separate regression. The vertical axis labels the dependent variable for each regression separately.

Psychological Outcomes. Furthermore, we evaluated the relationship between the proportion of peers who hold the gender-math stereotypical belief and three variables relating to students' beliefs: the likelihood that students will hold the same stereotypical belief, their perception of their parents' stereotypical beliefs, and their perception of peers' stereotypical beliefs. Figure 1 presents the results of estimating equation 1 with these three belief outcome variables as the dependent variable. We find that peer beliefs influenced the beliefs of boys and girls differently. Our findings suggest that a one *SD* increase in the proportion of peers who hold the stereotypical belief increased girls' likelihood of holding this belief themselves, girls' likelihood of perceiving their parents to hold this belief, and girls' likelihood of perceiving peers to have this belief, respectively, by 0.034 *SD*, 0.026 *SD*, 0.031 *SD* (p's < 0.1; see Appendix Table A.7).

We next examined whether the proportion of peers who believe that boys are better at learning mathematics than girls affects boys' and girls' perceptions of attention and praise from teachers in different ways. We find that having a greater proportion of peers who hold the gender-math stereotypical belief resulted in a significant decrease in girls' perception of attention ($b = -0.056$, $SE = .018$, $p < .01$) and praise from their math teachers ($b = -0.044$, $SE = .023$, $p < .1$), as compared to boys (see Figure 1). However, the proportion of peers who held gender-math stereotypes did not similarly influence boys' and girls' perceived attention and praise from English and Chinese teachers, although girls are somewhat less likely to perceive attention and praise from Chinese teachers (see Appendix Table A.8).

It is possible that girls exposed to more peers with the gender-math stereotypical belief may lower self-confidence and/or expectations for studying mathematics, which, in turn, will diminish their performance in math. Moreover, since math is one of the three core subjects in middle school, the degree of confidence a student has in studying math might also affect their perception of career choice in their future. It is also common for parents' expectations to be influenced by their children's academic performance. If exposure to peers who hold the gender-math stereotypical belief significantly influences students' academic performance, this may affect the degree of stress that students feel in relation to their parents' expectations about their academic achievement. Here we examine how gender-math stereotypical beliefs affect self-confidence and stress outcomes. Results suggest boys and girls have diverging estimates (see Appendix Table A.9). We found that exposure to more peers who believe Bm > Gm increased girls' likelihood of believing that math is difficult (b = -0.074, SE = .020, p < .01) and decreased their agreement that math relates to their future career advancement compared to boys (b = -0.054, SE = .021, p < .01). However, we did not find differential influence for boys and girls with respect to their perceptions of Chinese and English. This peer environment effect also contributed to higher levels of stress among girls about their parents' expectations regarding their academic achievement in comparison to their male counterparts (b = .075, SE = .028, p < .01).

The Laboratory Experiment

The natural experiment confirmed that a peer environment with high levels of the gender-math stereotype could hurt girls' math performance and reduce their pursuit in math-related extracurricular activities. Such peer environment also increases a child's likelihood of believing the gender-math stereotype, increases perceived difficulty of math and reduces aspirations among girls. However, the natural experiment does not directly test the mechanism. We assumed that a peer environment high in stereotypical peer beliefs might hurt girls' math performance due to the detrimental effect of the gender-math stereotype that might be transmitted through peers. But to establish a direct causal attribution of the stereotypical belief, we need to experimentally manipulate the activation of such belief and observe the immediate impact on female versus male students. Next, we conducted a laboratory experiment to explicitly activate gender-math stereotypes, followed by an assessment of test performance.

Method

The study was pre-registered at the Open Science Framework [https://osf.io/9wfsr/?view_only=6c2e525b828d4fb78dedc0d02bf4531c]. This study included a cross-sectional sample of undergraduate and graduate students in a large university in Southern China. A total of 547 undergraduate and graduate students were recruited through the behavioral economics laboratory on campus (167 men, 380 women; $Mean_{age}$ = 20.67 years, SD_{age} = 2.43, $Range_{age}$ = [18, 36]). The majority of the students majored in economics and had taken advanced college-level mathematics classes. Participation in the experiment was voluntary. All participants were compensated 30 RMB ($6.3 US dollars) as a standard participation fee and 3 RMB in cash for each question correctly answered[2]. This compensation rate was seen as attractive to an average student in China. Students were motivated to get high scores: the more correct answers they got, the more real bonus they would earn. Thus better performance in the lab was rewarded, just as better performance in the real world is rewarded.

In the experiment, 547 college students reported to the laboratory in mixed male and female groups of 19. Each participant has their own cubicle in the laboratory with minimal distraction. Participants were presented with a 5-minute video clip and answered questions regarding the content of the video. Participants were randomly assigned either to a *stereotype activation condition,* in which the video portrayed observable gender gaps in math performance in the United States and China, or to a *control condition* without gender or math-related information. The video in the *stereotype activation condition* contained statistics

[2] With the standard participation fee and piece rate bonus, the compensation per participant ranges from a minimum of 30 RMB to a maximum of 105 RMB.

from several representative surveys (e.g., China Family Panel Studies[3]) that suggest 1) men outperform women in mathematics on average, 2) more men than women score in the top percentiles in standardized math tests. The *stereotype activation* video is meant to trigger thoughts on gender group-based differences in math performance. It is intended to bring to the forefront any easy-to-activate stereotypical beliefs on gender math stereotypes. The video in the *control condition* talked about human memory and introduced strategies to facilitate memory encoding and strengthening, without gender differences mentioned. We expect the *control* video to evoke no gender or math-related thoughts for either women or men. In contrast, we hypothesized that the activated gender gap in the *stereotype activation condition* would evoke negative stereotypes and hamper math performance for female college students, but not for male college students.

After viewing the video and answering factual questions about the video content, participants performed a series of computer-based academic tests used to measure either math or verbal performance. They were informed that they would have 20 minutes to complete different types of test questions and that they would receive their scores at the end of the experiment. The 20 minutes were divided into four 5-minute test sessions[4]. Each of these test sessions was comprised of 5 advanced college-level math questions or 5 advanced college-level Chinese verbal questions. The order of the test sessions was counterbalanced. Math performance was calculated by the number of questions participants got correct out of all math questions (all students attempted all test questions). Similarly, verbal performance was calculated by the number of questions participants got correct out of all verbal questions. Upon completion of these test sessions, participants were offered a bonus test session where they could choose to complete either 5 more math questions or 5 more verbal questions. This

[3] The data are from China Family Panel Studies (CFPS), funded by the 985 Program of Peking University and carried out by the Institute of Social Science Survey of Peking University.

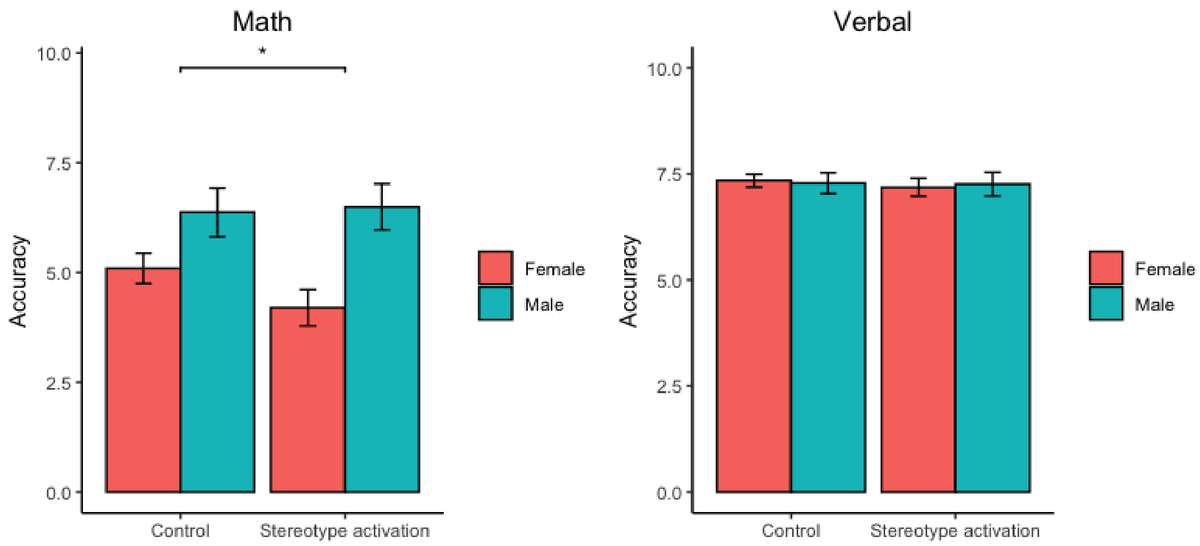
[4] Questions were adopted from Chinese college entrance exams. We selected moderately difficult math and verbal questions and piloted each test session to ensure that 5 minutes were the average completion time.

choice task was intended to measure their domain preference. We also collected participants' self-reported psychological outcomes, including test anxiety, domain identification, stereotype threat, confidence, effort, as well as general demographic information at the end of the experiment, but before they received test feedback on the number of correct answers and their compensation.

The Laboratory Experiment Results

Figure 2

Accuracy on the Math and Verbal Tasks in the Stereotype Activation and Control Conditions, for the Female and Male College Students.



Note: Math and verbal performance in the *stereotype activation condition* and the *control condition* for female college students and male college students in the laboratory experiment. Error bars reflect 95% confidence intervals. Top horizontal bars show two-way interaction (female versus male x stereotype activation versus control). *$p < 0.05$.

Math and verbal performance across gender groups in the laboratory experiment is plotted in Figure 2. In the control condition, we found an existing gender gap in math test scores between male and female students, which is consistent with the baseline gender-math gap in Chinese universities (we provide statistics regarding this realized gender gap in math in Chinese colleges by university rank using data from the 2010 to 2016 Chinese College Students Survey (CCSS) in the supplementary material). For male college students, they performed similarly well in math and verbal tasks across both the *control condition* and the *stereotype activation condition* (math: $t(164) = -0.31$, $p = .75$; verbal: $t(164) = .12$, $p = .90$). In contrast, in the *stereotype activation condition*, designed to trigger thoughts on gender-based differences in math, the female college students performed significantly worse in math tests compared with the female students in the *control condition* when no gender difference was elicited (math: $t(343) = 3.28$, $p = .0011$, Cohen's $d = 0.34$). No difference was found in verbal tests ($t(343) = 1.20$, $p = .23$). A two-way analysis of variance revealed a significant interaction between gender and condition on math performance ($F(1, 543) = 4.47$, $p = .035$). In other words, for both math and verbal tasks, the male college students were uninfluenced by condition, whereas the female college students performed significantly worse in math when the gender gap in math was made salient. Women in the *stereotype activation condition* scored 0.9 point lower on average ($M_{stereotype} = 4.19$, $SD = 2.70$; $M_{control} = 5.10$, $SD = 2.58$; $p = .001$), which represents an 18% drop in math performance, compared to those in the *control condition*. We conducted additional robustness checks (e.g., controlling for participant demographics and/or session fixed effects) in the supplementary material and found consistently significant results (see Appendix Table A.11). Contrary to the pre-registered hypothesis, we did not find a significant difference in students' domain preference—female and male college students were equally likely to choose to complete either a math bonus session or a verbal bonus session in the treatment versus control conditions ($p > .05$).

We hypothesized that the activation of the gender-math gap could potentially induce distracting thoughts among female students, who would be less able to exert effort in math tests and subsequently fare worse. Using the self-reported outcomes measured at the end of the experiment, we found a significant interaction effect between gender and condition on the ability to focus and exert effort in the tests. Female college students in the *stereotype activation condition* reported having more distracting thoughts and significantly less effort exerted in the tests ($M = 4.48$, $SD = 1.17$) compared to male students ($M = 4.88$, $SD = 1.15$, $p = 0.003$) or women in the *control condition* ($M = 4.78$, $SD = 1.11$, $p = 0.013$). We did not find a similar interaction effect on the other psychological outcomes measured, including academic domain identification, sense of threat, and test anxiety (p's > .05).

In sum, we found a robust effect of the activation of gender-math gaps on female college students' math performance. Female students in the *stereotype activation condition* exerted less effort and had more distracting thoughts during the experiment. The stereotype activation did not affect male college students and did not affect male or female students' verbal performance. However, these findings have limitations. The causal attribution of a laboratory study comes at the expense of some external validity. For example, we explicitly activated gender-math stereotypes in a laboratory environment. Such explicit priming may not mirror naturally occurring circumstances. In a natural environment, comments about gender-math stereotypes may come less explicitly and co-occur with social group interactions. The behavioral impact might follow much later after a series of encounters with such stereotypes, rather than observed immediately after one single encounter. It is possible that the gender-math gap puts pressure on boys to perform consistently well in math, creating a cognitive load comparable with that experienced by girls. It is also possible—though less plausible— that the girls structure their lives to avoid these concerns. The findings from the earlier natural experiment help address these issues and thus the two studies were complementary.

Discussion

Across two studies, the results advanced an important channel through which gendered beliefs in math ability are transmitted in one's ambient environment and affect female students' academic performance. We used data from a natural field experiment and a well-powered pre-registered laboratory experiment to demonstrate that gender-math stereotyping has both a cumulative and immediate influence on girls' and women's performance in mathematics. The age at which stereotyping brings to influence math performance starts much earlier before adulthood. A potential influence route is through peer environment: the more girls are exposed to peers who hold erroneous stereotypical beliefs on the innately superior math ability of boys, the worse they fare in math tests. We overcome the practical and ethical barriers to estimating this relationship experimentally by applying a quasi-experimental research design in a natural setting with substantial random variations in the proportion of a child's peers who hold the gendered belief in math ability.

Several aspects of the current research make it likely to underestimate the full impact of peer belief influence. First, we study one cohort of outcomes in middle schools (Study 1) and the momentary effect from a laboratory (Study 2), while the overall career effects of decades of cumulative exposure to these stereotypical beliefs are likely to be larger if they compound over time. Second, we study in contexts where female students are already strong in mathematics—in Study 1 middle school girls on average outperform boys and in Study 2 the majority of participants were college economics majors. Even in contexts when the stereotype runs counter to reality, we still find harm in exposure to peers who hold such beliefs.

While some suggest that the role of negative stereotype is "overcooked" and its influence on behavior is over-exaggerated (Jussim, 2015; Stoet & Geary, 2012), our research suggests the harm from gender-math stereotyping is still alive and perpetuating itself through peer environment. The prior research on gender inequality focuses almost exclusively on changing girls' attitudes and choices, with relatively less attention to the messages from peers with whom they share classrooms and schools on a daily basis (Riegle-Crumb & Morton, 2017). It may be more fruitful to change the ambient social environment that children and young adults are embedded in. We call attention to the importance of examining sources of peer influence within local contexts, as well as highlighting the need for more research that focuses on peer influence on stereotypical belief formation during the formative stages of adolescence in particular.

Our work also may provide a new perspective on the debate on stereotype threat. According to the stereotype threat theory, the possibility of confirming a negative stereotype for a target group provokes anxiety and threat, which leads group members to underperform on the task domain they are stereotyped on. Consistent with the main prediction, our laboratory experiment and field data support that negative stereotyping causally reduced female students' math performance. However, contrary to the increased anxiety or effort as proposed by classic stereotype threat theory, we found no evidence of changed levels of anxiety or threat. Instead, our data suggest that girls and young women reported significantly reduced effort (consistent with past results such as Mrazek et al., 2011; Jamieson & Harkins, 2007, 2009; Seitchik & Harkins, 2015), possibly from lowered self-confidence and self-expectations and increased distracting concerns (though our design cannot distinguish between a direct effect and a conditional effect). The prior theory also noted that targets of negative stereotypes must identify with the domain in which the threat occurs in order to be affected by a threat cue (Aronson et al., 1999; Deaux et al., 2007; Shih et al., 1999). The rationale is if targets do not care about the domain in which they are negatively stereotyped, then there should be no threat to the self to be concerned about and thus no stereotype threat effect. However, we did not find a moderating effect of domain identification or felt importance of mathematics—women and girls across the board seem to be affected by the negative stereotyping.

While our study addresses some of the limitations of prior research, it is subject to its own limitations. While we demonstrate the peer environmental influence on academic and psychological outcomes, we know less about the process of change from the belief transmission. While we capture the actual beliefs of the children and their peers, we do not have measures on what the children believe that their peers think as well as the actual classroom interactions and experiences. We speculate that the denser the gender-math stereotype is in a child's ambient peer environment, the more likely disparaging messages from peers will be transmitted through social interactions, leading to the observed gender inequality in aspiration and attainment. Future research could focus on the actual interactions that happen within peer groups.

These findings bear theoretical and practical implications. Social beliefs about differential abilities of gender groups contribute to individual belief formation (Bian et al., 2017; Jayachandran, 2015; Nollenberger et al., 2016;). We show that exposing a child to a greater number of peers who hold a stereotypical (but objectively inaccurate) belief causes a child to be more likely to hold the belief themselves. Moreover, the peer influence impacts not only belief formation but also demonstrated academic performance. Understanding the underlying mechanisms was critical for developing interventions that reduce stereotype threat and the group-based disparities that inspire this line of research. How to curb the peer transmission of gendered beliefs in ability? How to change these objectively inaccurate and proven harmful beliefs? Future research should aim to attend to theoretically driven interventions that mitigate and reverse the harm from beliefs in differential group-based ability.

Acknowledgments

We acknowledge support from National Natural Science Foundation of China (Grant Number: 72022015, 71988101).

References

Akabayashi, H., Nozaki, K., Yukawa, S., & Li, W. (2020). Gender differences in educational outcomes and the effect of family background: A comparative perspective from East Asia. *Chinese Journal of Sociology, 6*(2), 315-335.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29-46.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self‐efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development, 72*(1), 187-206.

Beutel, A. M., & Marini, M. M. (1995). Gender and values. *American Sociological Review*, *60*(3), 436-448.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children’s interests. *Science*, *355*(6323), 389-391.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica, 77*(4), 1003-1045.

Deaux, K., Bikmen, N., Gilkes, A., Ventuneac, A., Joseph, Y., Payne, Y. A., & Steele, C. M. (2007). Becoming American: Stereotype threat effects in Afro-Caribbean immigrant groups. *Social Psychology Quarterly, 70*(4), 384-404.

Eble, A., & Hu, F. (2022). Gendered beliefs about mathematics ability transmit across generations through children’s peers. *Nature Human Behaviour*, 1-12.

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*(1), 225-241.

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, *35*(2), 387-428.

Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality, 63*, 36-43.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25-44.

Fruehwirth, J. C., & Gagete-Miranda, J. (2019). Your peers' parents: Spillovers from parental education. *Economics of Education Review, 73*, 101910.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology*, *49*(10), 1886.

Gilligan, C. (1982). New maps of development: new visions of maturity. *American Journal of Orthopsychiatry, 52(*2), 199.

Gong, J., Lu, Y., & Song, H. (2018). The effect of teacher gender on students' academic and noncognitive outcomes. *Journal of Labor Economics*, *36*(3), 743-778.

Gu, X., & Jean Yeung, W. J. (2021). Why do Chinese adolescent girls outperform boys in achievement tests? *Chinese Journal of Sociology, 7*(2), 109-137.

Hu, F. (2015). Do girl peers improve your academic performance? *Economics Letters*, *137*, 54-58.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. Journal of Personality and Social Psychology, 93(4), 544–564.

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. Personality and Social Psychology Bulletin, 35(10), 1301–1314.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Economics*, *7*(1), 63-88.

Jussim, L. (2015, December). Is Stereotype Threat Overcooked, Overstated, and Oversold? Why dilute a great story with actual facts? *Psychology Today.* Retrieved from: https://www.psychologytoday.com/us/blog/rabble-rouser/201512/is-stereotype-threat-overcooked-overstated-and-oversold.

Lewis Jr, N., & Michalak, N. M. (2019). Has stereotype threat dissipated over time? A cross-temporal meta-analysis. *Preprint*.

Liu, R. (2018). Gender-math stereotype, biased self-assessment, and aspiration in STEM careers: The gender gap among early adolescents in China. *Comparative Education Review*, *62*(4), 522-541.

Mrazek, M. D., Chin, J. M., Schmader, T., Hartson, K. A., Smallwood, J., & Schooler, J. W. (2011). Threatened to distraction: Mind-wandering as a consequence of stereotype threat. Journal of Experimental Social Psychology, 47(6), 1243–1248.

Muralidharan, K., & Sheth, K. (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources, 51*(2), 269-297.

National Academy of Sciences. (2007). *Beyond bias and barriers:Fulfilling the potential of women in academic science and engineering* (Natl Acad Press, Washington, DC).

National Center for Science and Engineering Statistics NCSES (2015), “National Survey of College Graduates: Survey Description,” Public Use Documentation, available at https://ncsesdata.nsf.gov/datadownload/index.html?survey=PCG2015&typ=data (filename: NSCG_SD_2015.pdf).

Nollenberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, *106*(5), 257-61.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*(6), 1314.

OECD. (2015). *The ABC of Gender Equality in Education: Aptitude, Behavior, Confidence.* Paris: OECD.

Perry, B. L., Link, T., Boelter, C., & Leukefeld, C. (2012). Blinded to science: Gender differences in the effects of race, ethnicity, and socioeconomic status on academic and science attitudes among sixth graders. *Gender and Education, 24*(7), 725-743.

Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who Wants to Have a Career in Science or Math? Exploring Adolescents' Future Aspirations by Gender and Race/Ethnicity. *Science Education, 95*(3): 458–476.

Riegle-Crumb, C., & Morton, K. (2017). Gendered expectations: Examining how peers shape female students' intent to pursue STEM fields. *Frontiers in Psychology, 8*, 329.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?. In *Handbook of the Economics of Education* (Vol. 3, pp. 249-277). Elsevier.

Seitchik, A. E., & Harkins, S. G. (2015). Stereotype threat, mental arithmetic, and the mere effort account. Journal of Experimental Social Psychology, 61, 19–30.

Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*, *104*(12), 1514.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*(1), 80-83.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613.

Steele, C. M., & Vivaldi, W. (2010). How stereotypes affect us and what we can do. *W. W. Norton & Co.*

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, *16*(1), 93-102.

Wentzel, K. R. (2012). Teacher-student relationships and adolescent competence at school. In *Interpersonal Relationships in Education* (pp. 17-36). Brill.

Zigerell, L. J. (2017). Potential publication bias in the stereotype threat literature: Comment on Nguyen and Ryan (2008).