



# Stereotype Threat and Educational Tracking: A Field Experiment in Chinese Vocational High Schools

Socius: Sociological Research for a Dynamic World  
 Volume 4: 1–11  
 © The Author(s) 2018  
 Reprints and permissions:  
[sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav)  
 DOI: 10.1177/2378023118782011  
[srd.sagepub.com](http://srd.sagepub.com)

James Chu<sup>1</sup> , Prashant Loyalka<sup>1</sup>, Guirong Li<sup>2</sup>, Liya Gao<sup>2</sup>, and Yao Song<sup>2</sup>

## Abstract

Educational tracks create differential expectations of student ability, raising concerns that the negative stereotypes associated with lower tracks might threaten student performance. The authors test this concern by drawing on a field experiment enrolling 11,624 Chinese vocational high school students, half of whom were randomly primed about their tracks before taking technical skill and math exams. As in almost all countries, Chinese students are sorted between vocational and academic tracks, and vocational students are stereotyped as having poor academic abilities. Priming had no effect on technical skills and, contrary to hypotheses, modestly improved math performance. In exploring multiple interpretations, the authors highlight how vocational tracking may crystallize stereotypes but simultaneously diminishes stereotype threat by removing academic performance as a central measure of merit. Taken together, the study implies that reminding students about their vocational or academic identities is unlikely to further contribute to achievement gaps by educational track.

## Keywords

stereotype threat, social psychology, tracking, vocational education

Nearly all modern education systems sort and group students by ability (Baker 2014; Meyer and Ramirez 2009). The sorting of students into streams or tracks is often meant to help educators target learning needs more effectively, but educational tracks also generate differential expectations and stereotypes about students (Carbonaro 2005; Domina, Penner, and Penner 2017). Students in high-ability tracks are widely expected to perform well, whereas students in lower ability tracks are widely expected to perform poorly (Steenbergen-Hu, Makel, and Olszewski-Kubilius 2016).

If educational tracking generates stereotypes about students, do situations in which such stereotypes are made salient threaten students and undermine their performance? This question matters for two reasons. First, understanding the magnitude of stereotype threat effects in educational tracking informs debates regarding the costs and benefits of tracking (Heisig and Solga 2015; Lavrijsen and Nicaise 2016; Oakes 2005; Van de Werfhorst and Mijs 2010). Proponents of tracking generally argue that grouping students who have similar levels of ability helps teachers better address their needs (e.g., Gamoran and Mare 1989; Hallinan 1994). Critics counter that the stereotypes associated with

different ability groups are self-fulfilling (Eder 1981). Rather than improving learning for all students, tracking enhances educational inequalities by suppressing opportunities for certain students while enabling others to succeed. Understanding whether stereotype threat occurs in educational tracking would further clarify whether tracking improves learning for all students or benefits some students at the expense of others.

Second, investigating whether active stereotypes about educational tracks threaten performance helps further establish the theoretical boundaries of stereotype threat. Existing work on stereotype threat has largely relied on laboratory

<sup>1</sup>Stanford University, Stanford, CA, USA

<sup>2</sup>Henan University, Henan, China

## Corresponding Authors:

Guirong Li, Henan University, School of Education, Room 301, Jin Ming Avenue, Kaifeng, Henan 475004, China.  
 Email: [guirong1965@163.com](mailto:guirong1965@163.com)

James Chu, Stanford University, Department of Sociology, 450 Serra Mall, Building 120, Stanford, CA 94305, USA.  
 Email: [jchu1225@stanford.edu](mailto:jchu1225@stanford.edu)



studies to isolate how negative stereotypes about race, gender, and class interfere with performance (for a review of these studies, see Walton and Cohen 2003; Pennington et al. 2016). Investigating stereotype threat in educational tracks helps establish whether stereotype threat operates only for certain kinds of stereotypes (such as those for race, gender, and class) and, if so, why.

Related to the theoretical boundaries of stereotype threat is a question of real-world significance. Despite or perhaps because of long-standing scholarly interest in stereotype threat, there is competing evidence regarding whether stereotype threat explains real-world differences in group outcomes. Although some work suggests that stereotype threat explains real-world differences in achievement (Good, Aronson, and Inzlicht 2003; Huguot and Régner 2007; Walton and Spencer 2009), other work has raised questions about the straightforward application of stereotype threat explanations outside the laboratory setting (Cullen, Hardison, and Sackett 2004; Morgan, Mehta, and Research Library Core 2004; Müller and Rothermund 2014; Stricker and Ward 2004; Wax 2009; Wei 2012). At the very least, recent work has emphasized that the effects of stereotype threat by race depend on contextual moderators such as the racial composition of a given school (Hanselman et al. 2014) or local understandings of race (Herman 2009). Still other scholars point to possible publication biases in the literature on stereotype threat (Flore and Wicherts 2015; Nguyen and Ryan 2008; Stoet and Geary 2012; Zigerell 2017) and psychology more generally (Harris et al. 2013; Kahneman 2012). A well-powered empirical test for stereotype threat in a real-world setting would help inform this debate.

In this article, we examine whether activating negative stereotypes undermines student performance in the context of the most widespread form of educational tracking in the world today: vocational schooling. Almost all educational systems track students into academic or vocational education and training (VET) tracks at the secondary school level (Shavit and Müller 2000). For instance, a study across 28 countries, including contexts as diverse as the Czech Republic and Turkey, found that a median of 40 percent of school-aged students participate in upper secondary VET programs (Kuczera and Field 2010:32). Most students attend VET because they are unable to pass or do poorly on entrance examinations to attend academic high schools. Thus, although VET students may be seen as having technical skills, they are stereotyped as having low academic ability (Kuczera and Field 2010; Shavit and Müller 2000). Moreover, the fact that VET students are generally from poorer backgrounds compounds these negative stereotypes about academic ability. Policymakers hope that poor students can quickly attain skills before entering the labor market (Ainsworth and Roscigno 2005; Meer 2007), and thus VET schools tend to enroll students from economically poor families (Arum and Shavit 1995).

To investigate whether activating stereotypes about vocational tracks threatens the performance of VET students, we conducted a randomized controlled field study involving 115 Chinese VET schools and 11,624 first- and second-year students in the most popular majors at the time of the study (computing and digital control). Mirroring real-world test-taking conditions, students took fully proctored and standardized mathematics and technical skills tests linked to subject matter learned during the school year. At the beginning of each test, half of the students were randomly assigned to receive a prime in the form of a single question asking them to indicate whether they were in an academic or vocational track (following Steele and Aronson 1995). The other half of students were not exposed to this question.

In the sections that follow, we first describe vocational schooling in China to help establish its relevance as a context to study stereotype threat. Second, we review existing theories of stereotype threat to establish hypotheses to be tested in the experiment. Third, we describe how we implemented our field experiment and how we constructed key variables. Finally, we interpret the results from the experiment before concluding with implications for the literature at large.

### **Empirical Background: Stereotypes about Vocational Schooling in the Chinese Context**

Developing countries such as China have been investing heavily in VET programs. In the decade between 2001 and 2011, enrollment in Chinese VET increased from 11.7 million to 22.1 million students. Annual investments are now more than \$21 billion (Loyalka et al. 2015). Although the scale of enrollment and investment alone underlines the importance of studying this context, China is not alone in pushing forward with VET. Countries such as Brazil (Kuczera and Field 2010) and Indonesia (Newhouse and Suryadarma 2011) have passed legislation specifically identifying VET as a policy priority.

This interest in VET is motivated by a desire to reduce poverty and enhance economic development. Governments have explicitly cited poverty reduction as an animating reason for expanding vocational education (Kuczera and Field 2010). In principle, there are at least two reasons why VET is better suited to reducing poverty than academic schooling. First, academic high schools train students to enter college, but poor students are often faced with high opportunity costs. Each year they spend in school is a year of forgone wages. VET enhances earnings more directly by helping students develop skills and network with professionals (e.g., through internships and apprenticeships). Second, in most educational systems in the world, academic high schools are competitive. Low-income students often lack the resources or skills to achieve at a sufficient level to compete equally with higher income students. For this reason, poor students in

developing countries frequently stop attending school and instead enter the labor market. VET ostensibly fills this gap by diverting low-income students from the labor market and ensuring that they continue to learn skills that enhance their future wages.

Unfortunately, recent evidence suggests students are not learning much in VET. In a study of more than 10,000 students across two provinces of China, Loyalka et al. (2015) found that relative to attending academic high school, students in VET gain few technical skills and decline in their mathematics performance over time. Studies conducted in other countries, such as Romania and Indonesia, also suggest that VET students are learning little in school (Altinok 2012; Malamud and Pop-Eleches 2010). The fact that students appear to gain little from VET motivated our substantive interest in studying stereotype threat in this context. Given former studies showing how stereotype threat can suppress achievement in real-world contexts (e.g., Walton and Cohen 2007; Walton and Spencer 2009), we were concerned that VET students perform worse because of stereotype threat.

Indeed, prior work has demonstrated that Chinese VET students are negatively stereotyped as having poor academic performance (Yi et al. 2013). The most direct reason for negative stereotypes is the highly competitive Chinese education system. From an early age, individuals who perform well academically can expect to obtain higher educational tracks (Lam et al. 2004). At the end of junior high school and after nine years of compulsory education, students take a high school entrance examination to determine their tracks. Conditional upon taking the examination, students who do sufficiently well are given opportunities to continue onward in academic high schools (42 percent). Those who do not can choose to enter VET (25 percent). The remainder (33 percent) either enter the labor force or stay in junior high school an additional year to retake the exam (Song, Loyalka, and Wei 2013).

To be sure, poor performance on an exam is not usually sufficient to engender negative stereotypes. However, the broader cultural and historical context of China encourages interpretation of test performance as an index of character (Kipnis 2011), and performance on the exam thus becomes a salient tool for categorizing individuals (Woronov 2015). For instance, in a mixed vocational and academic secondary school in rural Zhejiang Province, parents asked principals to segregate vocational students away from academic students so that the VET students “would not infect the regular students” (Hansen 2015). The existence and salience of such stereotypes is also reflected in attempts to challenge them. In an ethnographic study of vocational schools, Woronov (2015) found that students frequently seek to challenge implications of flawed character that arise from poor test performance: “I actually did quite well on the high school entrance exam! But it would have been selfish of me to ask my parents to pay for a regular high school” (p. 51).

To further confirm that students in our sample recognized these negative stereotypes, we conducted a pilot study with a subsample of 687 VET students to answer a battery of questions about how other people generally see them. When asked to describe how others would judge their math ability, 89.2 percent of respondents reported that vocational school students were generally seen as having worse math abilities than academic track students.

## Theoretical Framework: Stereotype Threat

Stereotype threat occurs when individuals are at risk for confirming negative and widely shared expectations within a specific area or “domain” (Steele and Aronson 1995). When under stereotype threat, individuals demonstrate poorer performance in the negatively stereotyped domain (Steele 2010). For instance, African American students underperformed on a verbal test when asked to indicate their race before taking the test (study 4 in Steele and Aronson 1995). When a test was described as yielding gender differences (thus reminding test takers of the stereotype that women have weaker math ability), women demonstrated poorer math performance (Spencer, Steele, and Quinn 1999). When students of lower socioeconomic status were reminded of their marginalized status, they demonstrated weaker performance on a verbal test (Croizet and Claire 1998).

Stereotype threat is thought to be activated by situational cues (henceforth “primes”) that remind individuals that their group is negatively stereotyped (Aronson et al. 1999). These reminders lead to negative performance via several mechanisms (Pennington et al. 2016; Schmader and Johns 2003), such as depleting working memory (Schmader, Johns, and Forbes 2008) or increasing mental load (Croizet et al. 2004). For instance, Asian American women tend to do worse in mathematics when their gender identity is primed but not when their Asian identity is primed (Shih, Pittinsky, and Ambady 1999). Even subtle primes can generate stereotype threat (Danaher and Crandall 2008; Nguyen and Ryan 2008; Steele, Aronson, and Spencer 2007). As noted above, merely indicating one’s race or gender before a test can trigger stereotype threat (Steele and Aronson 1995). Simply being a numerical minority in a room appears to trigger stereotype threat among women (Murphy, Steele, and Gross 2007).

Scholars have identified several important conditions that must be present before stereotypes are threatening (Cesario 2014; Lamont, Swift, and Abrams 2015; Van Bavel et al. 2016). Some conditions are straightforward: an individual must recognize a widely shared expectation of poor performance on a particular domain (Steele 2010). Notably, even if individuals do not personally agree with the stereotypes, individuals are still threatened because their abilities are questioned as a result of their group identity. Another condition is that the test for performance be sufficiently challenging. From

a methodological perspective, an easy task means that all students perform relatively well, leading to reduced variance and statistical problems of censoring. From a theoretical perspective, a test that is extremely easy nullifies threats by negative stereotypes (Keller 2007).

A third condition is that the test must be construed as diagnostic of ability. Spencer et al. (1999) showed that women who were told that math performance on a given test indicated differences in ability performed worse than women who were not told that the test was ability diagnostic. Among the most important scope conditions is domain identification: individuals are threatened only if they see the subject or area being tested as important to them (Aronson et al. 1999; Spencer et al. 1999; Steele et al. 2007). If a student negatively stereotyped as bad at mathematics does not care about doing well in mathematics, that student has no reason to feel threatened (Penner and Willer 2011). As Steele et al. (2007) noted, “one ironic and unfortunate aspect” of stereotype priming is that those who care the most about a given domain are those that will underperform.

Taking the literature and context together, we propose the following hypotheses. First, given sufficiently difficult tests that are framed as diagnostic of ability, reminding students about their identity as vocational school students will trigger negative stereotypes about their academic abilities and, in turn, undermine their performance on academic but not technical subjects. Indeed, we hypothesize that they may even perform better on their technical subjects. Second, any stereotype threat effects should be strongly moderated by domain identification. Students must think that a particular subject is important before they feel threatened, and students who do not identify with a certain subject will not experience stereotype threat (or its effects will be greatly diminished).

## Data and Methods

### Sample and Experimental Setup

To test these hypotheses, we conducted an experiment among a sample of VET schools in a province in central China. The experiment was conducted as part of a broader study to compare the learning outcomes of students enrolled in academic versus vocational tracks. This province is of average income: of the 31 provinces in China, this province ranks 22nd in terms of gross domestic product per capita; although it is becoming increasingly urbanized, 55 percent of the population lives in rural areas. The province is also among the largest in China in terms of population.

We restricted our sampling frame to schools that had at least one of the two most popular majors in this province: computing and digital control.<sup>1</sup> This sampling restriction was imposed because each major has a different curriculum,

and we needed to create tailored examinations for each major. We further restricted the sample to schools with enrollments over 30, because schools with fewer than 30 students were likely to close before our enumerators could administer the examinations. A total of 115 schools (among a universe of approximately 600 schools in the province) had a computing or digital control major and had first-year enrollments over 30.

In each school, we randomly sampled one first-year class and one second-year class in the computing or digital control major. Of these 115 schools, 102 had a computing major and 63 had a digital control major, yielding a total of 328 classes sampled ( $[115 + 63] \times 2$ ). There were 7,874 first-year and 5,249 second-year students across these 328 classes ( $n = 13,123$ , or approximately 40 students per class). No incentives were given as part of the study, and approximately 10 percent of participants declined to participate in the study, leaving us with 11,624 students. The sample represents VET students in the two most popular majors in one central province of China.

Each student in the sample was asked to take math and technical skills (either in computing or digital control) examinations on the basis of national curricular standards. Students were first told that this exam was a test of their abilities as part of a broader project to compare students across high schools in the province. Enumerators did not specify whether this referred only to VET or both academic and VET schools, allowing the prime to highlight the comparison between tracks. Students were then instructed that they had 30 minutes to complete each test and that all tests were fully proctored.

The stereotype prime was implemented as follows. Before the timed examination began, students were asked to write down their names and the names of their schools on a cover sheet of the examination. Half of the students were randomly assigned to respond to an additional question on the cover sheet: “What educational track are you currently enrolled in—vocational or academic?” The other half did not receive this question. To ensure that the students noticed this question, enumerators instructed students to double-check their cover pages for any blanks before beginning the test. Despite this protocol, 93 students (0.80 percent of the total of 11,624) had missing information on their cover pages. Because we could not be sure that they noticed the prime, we drop these 93 students from our analysis below (total  $n = 11,531$ ).

After the prime was established, the vocational exam was administered first. If students received the prime on their vocational exam, they would also receive it on their math exam. All enumerators were blind to treatment assignment: the primes were printed directly on randomly assigned examinations, and enumerators passed out examinations face down. Finally, random assignment occurred within each classroom. Because assignment is random only conditional on classroom, the results below must use regressions with classroom fixed effects to properly compare the average

<sup>1</sup>Digital control prepares students to operate computerized machinery in factories.

treatment outcome of students who received the prime versus those who did not. Finally, because the sample for our study is clustered by classrooms, our standard errors are adjusted for clustering at the classroom level ( $n = 328$ ).

We acknowledge limitations to this research design. For instance, because the prime tested is subtle, any effect sizes measured are conservative or potentially even a lower bound (e.g., Müller and Rothermund 2014). Nonetheless, the choice of using a subtle test of stereotype threat was deliberate. In real-world exams, students are rarely faced with clear primes that follow laboratory conditions. By contrast, this study tests a prime that is more likely to occur in real-world conditions: students are simply asked to report if they belong to a VET or academic track.

Indeed, the major strengths of this research design are its statistical power and ecological validity. To our knowledge, this is the largest experimental test of stereotype threat to date. We test both math and technical skills, which allows us to better identify if the effects are domain specific. Moreover, the study tests whether stereotype threat reduces the performance of Chinese vocational high school students on real-world math or technical skills tests.

### *Measures and Data Collection*

The dependent variable was performance on a math or technical skills test. Because we wanted the results to reflect how stereotype threat might realistically affect student performance, we sought to administer tests that could capture what students were learning in school. Thus, these standardized examinations were developed in consultation with national curricular standards for vocational schooling. Specifically, we sourced items (questions) from Chinese national math, computer, and digital control skill tests. In addition, we piloted these items with a panel of vocational school students who were not in our sample, using their responses to ensure that the questions were indeed capturing students' math and vocational achievement and also to optimize psychometric properties. One important change we made as a result of the pilot was to add nine difficult questions to the second-year computing exam, as the test was too easy. We used these results to develop six standardized examinations of adequate difficulty: math, computing, and digital control for first- and second-year students, respectively (see Figure 1 and Supplementary Information for details regarding the distribution of test scores).

After students completed the two exams (i.e., after the experiment was completed), we administered a questionnaire with a series of questions about student age, gender, ethnicity, number of individuals in their households, and type of household registration. (Household registration in China refers to whether a student is categorized as urban or rural on his or her identification.) To control for student background, we asked for parental education and whether their parents primarily pursued farm work. One of the main reason for constructing these covariates is to check for balance in

treatment assignment. Specifically, Appendix Table A1 shows results from a regression of eight key covariates on treatment assignment. This implies that the treatment and control groups are properly balanced and that randomization was successful. In addition to checking for balance, we use these covariates as control variables to improve statistical power. However, we present results both including and omitting such variables to show that the results are robust to our choice of covariates.

From the questionnaire, we also construct two measures for identification with the academic domain. The first measure is coded from a question asking students to identify the kind of degree they aspire to attain: "In an ideal world, what is the highest degree you would wish to obtain?" Student choices were vocational high school (8.3 percent of respondents), two-year vocational college (35.8 percent), four-year academic college (37.8 percent), above academic college (15.5 percent), and other (2.5 percent). We coded students who aspired to attend academic colleges or above as identifying with the academic domain.

As a further robustness check, we also coded for identification with the academic domain by asking students to report on their intentions. After all, not all individuals who aspire to the academic track are being realistic, and thus not all students plan to act on their aspirations. This second measure is coded from the following question: "What do you plan to do after leaving vocational high school?" Students who responded that they intended pursue academic college (14.1 percent) were coded as identifying with the academic domain. The reference category is all other intentions, such as finding a job (34.1 percent) or starting a company (6.2 percent). Summary statistics for all measures used in the study can be found in Table 1.

## **Results**

As this is an experiment, the mean differences across treatment groups provides a first-order estimate of the treatment effect. Figure 2 summarizes these raw means by the primed versus nonprimed groups for both the technical skill and math outcomes. The plot includes 95 percent confidence intervals to help the reader assess statistical significance. The main result that emerges from Figure 2 is that there is no statistically significant effect of the prime on technical skills. The prime, however, does appear to slightly improve students' math performance. This is the opposite of our hypothesis, implying that the students performed better when reminded about their vocational school identity.

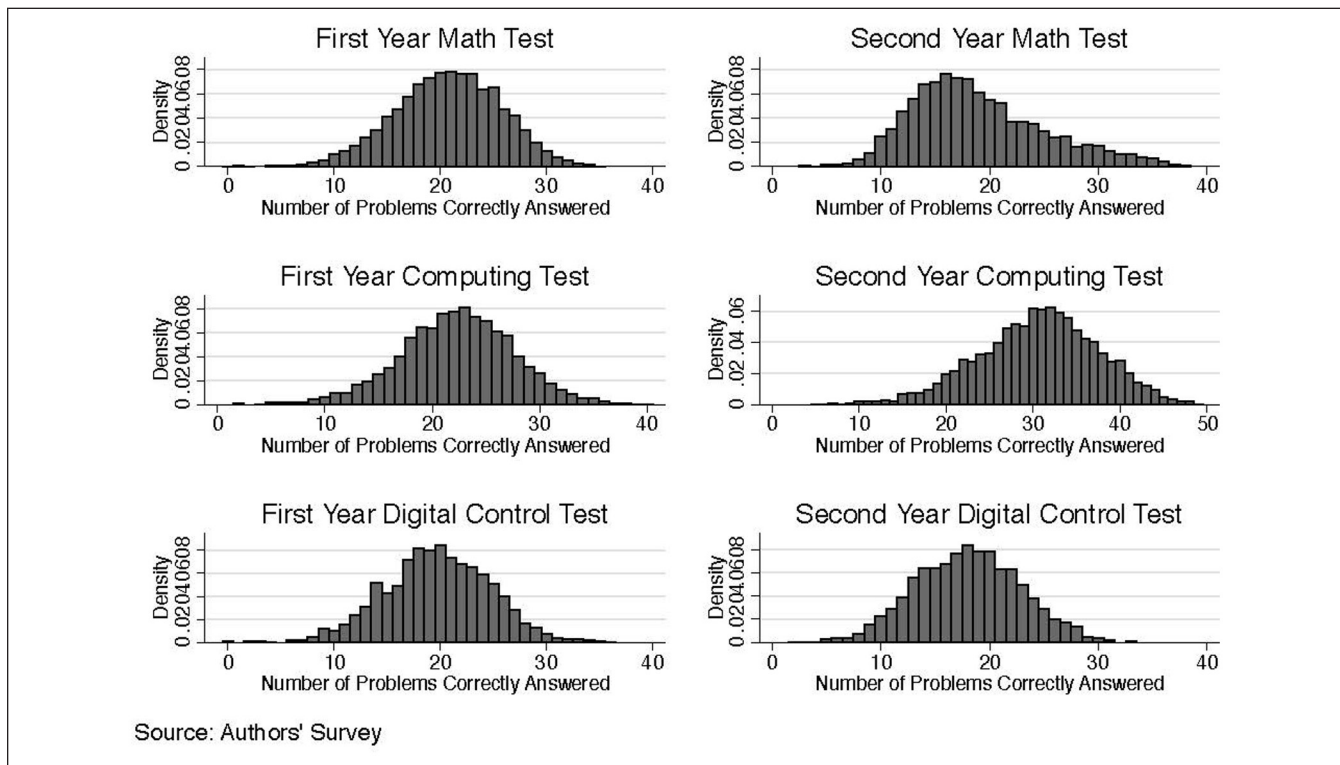
As noted earlier, the inclusion of classroom fixed effects is required for unbiased results. After including classroom fixed effects and adjusting standard errors for clustering at the classroom level, the results confirm that being subtly reminded of one's marginalized academic status appears to increase math performance. To our surprise, students who received the prime performed 0.032 standard deviations

**Table 1.** Descriptive Statistics of Variables Used in the Study.

	Mean	SD	Minimum	Maximum
Treatment indicator				
Stereotype primed (1 = yes)	0.501	0.500	0	1
Outcomes				
Standardized math score	0	1	-4.142	3.144
Standardized technical skills score	0	1	-4.013	3.504
Covariates				
First-year student (1 = yes)	0.589	0.492	0	1
Age (years)	17.87	1.342	14	24
Male (1 = yes)	0.725	0.447	0	1
Has urban household registration (1 = yes)	0.124	0.329	0	1
Number of members of household	4.628	1.138	1	10
Parents do not have junior high diploma (1 = yes)	0.237	0.425	0	1
Parents are primarily farmers (1 = yes)	0.300	0.458	0	1
Domain identification				
Aspire to attain an academic degree	0.533	0.499	0	1
Intend to attain an academic degree	0.142	0.349	0	1

Source: Authors' survey.

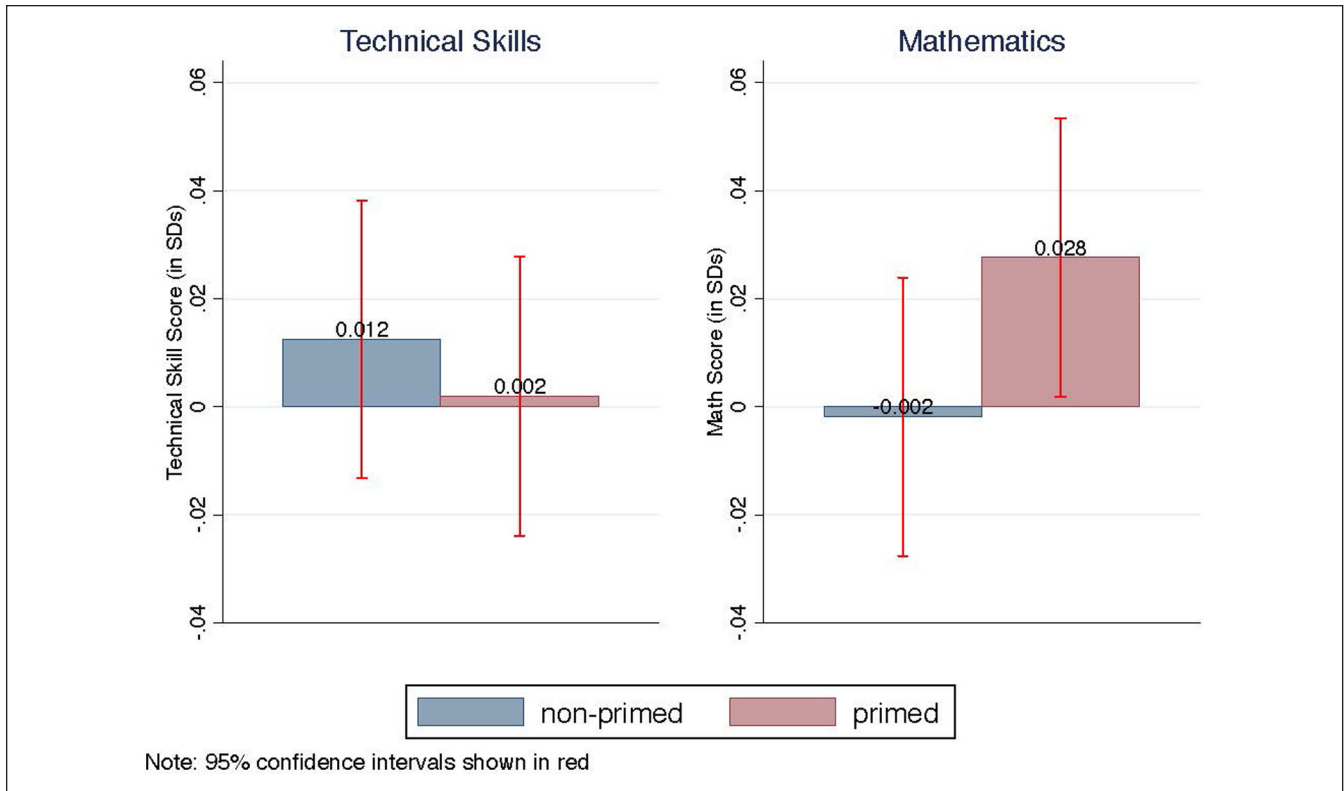
Note:  $N = 11,531$  (93 students who failed a manipulation check were excluded).

**Figure 1.** Distribution of test scores across exams.

better ( $p = .067$ ; Table 2, column 1). Put in terms of problems solved, students solved 0.17 more problems when primed about their VET identity. By contrast, there appears to be little evidence to suggest that the prime changed students' performance on the technical skills test (column 3). The

results are substantively identical when adjusted for covariates (columns 2 and 4).

As noted above, one of the scope conditions of stereotype threat is domain identification, and we hypothesized that stereotypes would selectively threaten students identified with



**Figure 2.** Raw means across math and technical skills outcomes.

**Table 2.** Effect of Stereotype Prime on Students’ Test Performance (Standardized).

	(1)	(2)	(3)	(4)
	Mathematics		Technical Skills	
Stereotype prime (1 = yes)	0.032* (0.017)	0.032* (0.017)	-0.004 (0.016)	-0.005 (0.016)
Age		0.050*** (0.009)		0.045*** (0.008)
Male (1 = yes)		-0.141*** (0.026)		0.062** (0.028)
Han ethnicity (1 = yes)		-0.018 (0.066)		-0.012 (0.075)
Number of members of household		-0.000 (0.008)		-0.016* (0.008)
Has urban household registration (1 = yes)		-0.157*** (0.027)		0.037 (0.033)
Parents lack junior high education (1 = yes)		0.000 (0.021)		-0.002 (0.021)
Parents are farmers (1 = yes)		0.015 (0.018)		-0.026 (0.021)
Constant	-0.003 (0.009)	-0.756*** (0.186)	0.009 (0.008)	-0.741*** (0.171)
R <sup>2</sup>	.241	.250	.211	.215

Note: Standard errors adjusted for clustering at the classroom level are shown in parentheses. All results include fixed effects for 328 classrooms in the study.  $N = 11,531$ .

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

the academic domain. Table 3 shows that the average effects indeed differ by domain identification, but the effects are the opposite of what we expected. As expected, students who aspire to attain academic degrees perform 0.18 standard deviations better in math than those who do not (column 1, row 2). To our surprise, priming the VET identity of students who do not aspire to attain academic degrees improves their math scores by 0.06 standard (column 1, row 1;  $p = .011$ ). This effect declines by 0.06 standard deviations, however,

for primed students who also identify with the academic domain (column 1, row 3;  $p = .042$ ). This implies that the total effect for students identifying with the academic domain is zero but positive for students who do not identify with the academic domain. This pattern of results again is the opposite of what we hypothesized.

The same pattern of effects holds if we measure identification by actual intentions to pursue academic degrees. As expected, students who intend to pursue academic degrees

**Table 3.** Domain Identification Moderates Effect of Priming on Standardized Student Performance.

	(1)	(2)	(3)	(4)
	Mathematics		Technical Skills	
Stereotype prime (1 = yes)	0.06*** (0.02)	0.05** (0.02)	0.00 (0.02)	-0.00 (0.02)
Aspire to attain academic degree (1 = yes)	0.18*** (0.02)		0.14*** (0.03)	
Aspire × Primed	-0.06** (0.03)		-0.01 (0.03)	
Intend to pursue academic degree (1 = yes)		0.45*** (0.04)		0.29*** (0.04)
Intend × Primed		-0.09* (0.05)		-0.01 (0.05)
Covariates adjusted	No	No	No	No
Constant	-0.10*** (0.02)	-0.07*** (0.01)	-0.06*** (0.02)	-0.03*** (0.01)
R <sup>2</sup>	.246	.258	.215	.219

Note: Standard errors adjusted for clustering at the classroom level are shown in parentheses. All results include fixed effects for 328 classrooms in the study. N = 11,531 (53.3 percent of students aspired to an academic degree, and 14.17 percent intended to pursue one).

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

perform 0.45 standard deviations better than their peers who desire to go to work or pursue additional training (column 2, row 4). Among students without intentions to pursue academic schooling, the prime increases math performance by 0.05 standard deviations (column 2, row 1;  $p = .013$ ). However, the effect declines by 0.09 standard deviations among students who intend to pursue academic degrees (column 2, row 5;  $p = .052$ ). Although this means the net effect does undermine math performance among domain identified students by 0.04 standard deviations (0.09–0.05), this net effect is not statistically significant ( $p = .312$ ).

Finally, domain identification in math does not lead to heterogeneous effects in technical skills: the domains are independent. Columns 3 and 4 in Table 3 show that there is neither a main (row 1) nor an interaction effect of the prime (rows 3 and 5). This means that students who identify with the academic domain do not experience different priming effects. Moreover, these results are substantively identical even when covariates are included (see Appendix Table A2).

## Discussion and Conclusion

Educational tracking is a part of almost all modern societies; to what extent do situations in which such stereotypes are made salient threaten students and undermine their performance? Our study finds that priming students about their VET identity appears to have no effect on technical or math performance. If anything, priming leads to a slight improvement in math performance on average. This overall effect can be decomposed by domain identification. Students who either aspire or intend to pursue academic degrees are neither helped nor harmed by priming, while students who identify with the vocational domain perform up to 0.06 standard deviations better on math examinations when reminded about their VET identity.

In interpreting these results, we first underline that the effects are substantively modest in magnitude. Statistical

significance does not equate to substantive significance, and even small effects can be detected with the large sample size used in this study. To provide a point of comparison, the largest effects measured here (0.06 standard deviations) are less than the smallest effects summarized in the literature (0.11 standard deviations, among women who do not identify as being good at math). The smallest effects in the literature are approximately twice the magnitude of the effects observed here (Nguyen and Ryan 2008).

What is clear is that the results uniformly fail to support our hypotheses. We do not find evidence supporting the claim that activating negative stereotypes about VET undermines math performance. There is also no evidence that stereotype threat occurs primarily among students who care about academic subjects such as mathematics. Instead and to our surprise, the interaction effects indicate a boost in performance among students who do not identify with the academic domain.

How do we explain these largely null results? Across several robustness checks, we rule out interpretations that students were not paying attention to the prime, that the stereotype prime wore off, that students needed to be seated next to academic high school students to feel threatened, that students needed to identify more strongly as vocational school students, or that the tests were not difficult enough (see Supplementary Information for details of these tests). However, this does not enable us to rule out stereotype threat along educational tracks: it is plausible that students were already under some form of stereotype threat and the effect of the prime was nonexistent because the threat was already present. Instead, what is clear across all these interpretations is that student performance is unlikely to be undermined via the kind of primes tested in this study. Especially noteworthy is that this prime is a real-world practice that we believed could plausibly trigger stereotype threat: these are customary questions students actually receive when taking high-stakes exams for college, scholarships, and other opportunities.



Why might the prime have led to a slight improvement in math performance? To account for improvement in math performance, we provide preliminary evidence showing that priming students about their vocational identity reminded them that they were no longer being judged solely on their academic performance (see Supplementary Information). Although we recognize the substantive effect size of the boost in math performance is minuscule, this surprising reversal of what we expected suggests that stereotype threat along educational track may differ from race, gender, or class. For instance, educational tracking necessarily sorts students by academic ability and thus crystallizes certain negative stereotypes. Although we hypothesized that the negative stereotypes would threaten student performance, perhaps VET also gives students who are either unable to or uninterested in competing in highly competitive academic tracks an opportunity to develop alternative, technical skills. By removing academic performance as a core measure of merit, VET creates a new reference group and helps students avoid comparison with academically tracked students. This dynamic is generally not present for race, class, or gender, because individuals generally cannot easily change identification along these lines.

Because the study was conducted on a random sample of vocational school students in one province of China, the results generalize to the approximately 2 million VET students enrolled in this province. How do these results generalize outside of this context and to other forms of tracking? Given our exploratory analyses, it is important to stress that educational tracking does not always move students to a new basis of merit. Within-school educational tracking may divide students into regular and advanced courses, creating negative expectations and indeed threaten student performance. We would not expect the results from our setting to hold in such contexts. Instead, the results generalize best to contexts with established VET programs and clear, negative stereotypes about the academic ability of VET students. In this way, the Chinese educational context resembles those in countries such as Romania and Indonesia (Altinok 2012; Malamud and Pop-Eleches 2010). These are countries that have rapidly expanding VET programs, competitive entrance examinations, and traditionally negative stereotypes about the academic abilities of VET students.

At a methodological level, this study answers recent calls to replicate experimental studies and assess their real world significance (Kahneman 2012; Simmons, Nelson, and Simonsohn 2011). The current stereotype threat literature is criticized for its reliance on multiple experiments with small samples, which are known to increase the likelihood of false positives (Button et al. 2013). The major strength of this study is its statistical power and ecological validity. The size of the experiment makes it the highest powered test of stereotype threat (to our knowledge). Moreover, students were taking tests that were pegged to topics they were learning on a day to day basis, and the stereotype prime—a question

asking students simply to indicate whether they are in a VET or academic track—could occur in real settings. Although our results were not what we expected, they highlight the importance of future experimental studies that test whether reducing stereotype threat improves student performance.

Ultimately, for scholars interested in establishing the costs and benefits of tracking, the results suggest that stereotypes are unlikely to hamper student performance, at least via triggers of vocational identity on examinations. We expected negative stereotypes about academic ability to undermine the math performance of VET students. In contrast, the results suggest that priming VET students about their track has no substantive effects (and perhaps even modestly enhances student performance).

### Acknowledgments

We thank the Stanford Inequality Workshop, David Pedulla, Florencia Torche, and Robb Willer for helpful comments in preparation of this article. James Chu was supported by the National Science Foundation Graduate Research Fellowship under grant DGE-114747. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

### ORCID iD

James Chu  <https://orcid.org/0000-0003-2702-470X>

### References

- Ainsworth, James W., and Vincent J. Roscigno. 2005. "Stratification, School-work Linkages and Vocational Education." *Social Forces* 84(1):257–84.
- Altinok, Nadir. 2012. "General versus Vocational Education: Some New Evidence from PISA 2009." Technical Report. Paris, France: UNESCO.
- Aronson, Joshua, Michael J. Lustina, Catherine Good, Kelli Keough, Claude M. Steele, and Joseph Brown. 1999. "When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat." *Journal of Experimental Social Psychology* 35(1):29–46.
- Arum, Richard, and Yossi Shavit. 1995. "Secondary Vocational Education and the Transition from School to Work." *Sociology of Education* 68(3):187–204.
- Baker, David. 2014. *The Schooled Society: The Educational Transformation of Global Culture*. Stanford, CA: Stanford University Press.
- Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14(5):365–76.
- Carbonaro, William. 2005. "Tracking, Student Effort, and Academic Achievement." *Sociology of Education* 78(1):27–49.
- Cesario, Joseph. 2014. "Priming, Replication, and the Hardest Science." *Perspectives on Psychological Science* 9(1):40–48.

- Croizet, Jean-Claude, and Theresa Claire. 1998. "Extending the Concept of Stereotype Threat to Social Class: The Intellectual Underperformance of Students from Low Socioeconomic Backgrounds." *Personality and Social Psychology Bulletin* 24(6):588–94.
- Croizet, Jean-Claude, Gérard Després, Marie-Eve Gauzins, Pascal Huguet, Jacques-Philippe Leyens, and Alain Méot. 2004. "Stereotype Threat Undermines Intellectual Performance by Triggering a Disruptive Mental Load." *Personality and Social Psychology Bulletin* 30(6):721–31.
- Cullen, Michael J., Chaitra M. Hardison, and Paul R. Sackett. 2004. "Using SAT-grade and Ability-job Performance Relationships to Test Predictions Derived from Stereotype Threat Theory." *Journal of Applied Psychology* 89(2):220–30.
- Danaher, Kelly, and Christian S. Crandall. 2008. "Stereotype Threat in Applied Settings Re-examined." *Journal of Applied Social Psychology* 38(6):1639–55.
- Domina, Thurston, Andrew Penner, and Emily Penner. 2017. "Categorical Inequality: Schools as Sorting Machines." *Annual Review of Sociology* 43(1):311–30.
- Eder, Donna. 1981. "Ability Grouping as a Self-fulfilling Prophecy: A Micro-analysis of Teacher-student Interaction." *Sociology of Education* 54(3):151.
- Flore, Paulette C., and Jelte M. Wicherts. 2015. "Does Stereotype Threat Influence Performance of Girls in Stereotyped Domains? A Meta-analysis." *Journal of School Psychology* 53(1):25–44.
- Gamoran, Adam, and Robert D. Mare. 1989. "Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?" *American Journal of Sociology* 94(5):1146–83.
- Good, Catherine, Joshua Aronson, and Michael Inzlicht. 2003. "Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat." *Journal of Applied Developmental Psychology* 24(6):645–62.
- Hallinan, Maureen T. 1994. "Tracking: From Theory to Practice." *Sociology of Education* 67(2):79–91.
- Hanselman, Paul, Sarah K. Bruch, Adam Gamoran, Geoffrey D. Borman, and Sage Paul Hanselman. 2014. "Threat in Context: School Moderation of the Impact of Social Identity Threat on Racial/Ethnic Achievement Gaps." *Sociology of Education* 87(2):106–24.
- Hansen, Mette Halskov. 2015. *Educating the Chinese Individual: Life in a Rural Boarding School*. Seattle: University of Washington Press.
- Harris, Christine R., Noriko Coburn, Doug Rohrer, and Harold Pashler. 2013. "Two Failures to Replicate High-performance-goal Priming Effects." *PLoS ONE* 8(8):e72467.
- Heisig, Jan Paul, and Heike Solga. 2015. "Secondary Education Systems and the General Skills of Less- and Intermediate-educated Adults." *Sociology of Education* 88(3):202–25.
- Herman, Melissa R. 2009. "The Black-white-other Achievement Gap: Testing Theories of Academic Performance among Multiracial and Monoracial Adolescents." *Sociology of Education* 82(1):20–46.
- Huguet, Pascal, and Isabelle Régner. 2007. "Stereotype Threat among Schoolgirls in Quasi-ordinary Classroom Circumstances." *Journal of Educational Psychology* 99(3):545–60.
- Kahneman, Daniel. 2012. "A Proposal to Deal with Questions about Priming Effects." *Nature* 490:2.
- Keller, Johannes. 2007. "Stereotype Threat in Classroom Settings: The Interactive Effect of Domain Identification, Task Difficulty and Stereotype Threat on Female Students' Maths Performance." *British Journal of Educational Psychology* 77(2):323–38.
- Kipnis, Andrew B. 2011. *Governing Educational Desire: Culture, Politics, and Schooling in China*. Chicago: University of Chicago Press.
- Kuczera, Malgorzata, and Simon Field. 2010. "OECD Reviews of Vocational Education and Training: A Learning for Jobs Review of Germany." Technical Report.
- Lam, Shui-Fong, Pui-Shan Yim, Josephine S. F. Law, and Rebecca W. Y. Cheung. 2004. "The Effects of Competition on Achievement Motivation in Chinese Classrooms." *British Journal of Educational Psychology* 74(2):281–96.
- Lamont, Ruth, Hannah Swift, and Dominic Abrams. 2015. "A Review and Meta-analysis of Age-based Stereotype Threat: Negative Stereotypes, Not Facts, Do the Damage." *Psychology and Aging* 30(1):180–93.
- Lavrijssen, J., and I. Nicaise. 2016. "Educational Tracking, Inequality and Performance: New Evidence from a Differences-in-differences Technique." *Research in Comparative and International Education* 11(3):334–49.
- Loyalka, Prashant, Xiaoting Huang, Linxiu Zhang, Jianguo Wei, Hongmei Yi, Yingquan Song, Yaojiang Shi, and James Chu. 2015. "The Impact of Vocational Schooling on Human Capital Development in Developing Countries: Evidence from China." *World Bank Economic Review* 30(1):lhv050.
- Malamud, Ofer, and Cristian Pop-Eleches. 2010. "General Education versus Vocational Training: Evidence from an Economy in Transition." *Review of Economics and Statistics* 92(1):43–60.
- Meer, Jonathan. 2007. "Evidence on the Returns to Secondary Vocational Education." *Economics of Education Review* 26(5):559–73.
- Meyer, John W., and Francisco O. Ramirez. 2009. "The World Institutionalization of Education." Pp. 111–32 in *Discourse Formation in Comparative Education*, edited by Jurgen Schriewer. New York: Peter Lang.
- Morgan, Stephen L., Jal D. Mehta, and Research Library Core. 2004. "Beyond the Laboratory: Evaluating the Survey Evidence for the Disidentification Explanation of Black-white Differences in Achievement." *Library* 77(1):82–101.
- Müller, Florian, and Klaus Rothermund. 2014. "What Does It Take to Activate Stereotypes? Simple Primes Don't Seem Enough: A Replication of Stereotype Activation." *Social Psychology* 45(3):187–93.
- Murphy, Mary C., Claude M. Steele, and James J. Gross. 2007. "Signaling Threat: How Situational Cues Affect Women in Math, Science, and Engineering Settings." *Psychological Science* 18(10):879–885.
- Newhouse, David, and Daniel Suryadarma. 2011. "The Value of Vocational Education: High School Type and Labor Market Outcomes in Indonesia." *World Bank Economic Review* 25(2):296–322.
- Nguyen, Hannah-Hanh D., and Ann Marie Ryan. 2008. "Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-analysis of Experimental Evidence." *Journal of Applied Psychology* 93(6):1314–34.

- Oakes, J. 2005. *Keeping Track: How Schools Structure Inequality*. 2nd ed. New Haven, CT: Yale University Press.
- Penner, Andrew M., and Robb Willer. 2011. "Stigma and Glucose Levels: Testing Ego Depletion and Arousal Explanations of Stereotype Threat Effects." *Current Research in Social Psychology* 16(3):n3.
- Pennington, Charlotte R., Derek Heim, Andrew R. Levy, and Derek T. Larkin. 2016. "Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators." *PLoS ONE* 11(1):e0146487.
- Schmader, Toni, and Michael Johns. 2003. "Converging evidence That Stereotype Threat Reduces Working Memory Capacity." *Journal of Personality and Social Psychology* 85(3):440–52.
- Schmader, Toni, Michael Johns, and Chad Forbes. 2008. "An Integrated Process Model of Stereotype Threat Effects on Performance." *Psychological Review* 115(2):336–56.
- Shavit, Yossi, and Walter Müller. 2000. "Vocational Secondary Education, Tracking, and Social Stratification." Pp. 437–53 in *Handbook of the Sociology of Education*, edited by M. Hallinan. New York: Springer.
- Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady. 1999. "Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance." *Psychological Science* 10(1):80–83.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22(11):1359–66.
- Song, Yingquan, Prashant Loyalka, and Jianguo Wei. 2013. "Determinants of Tracking Intentions, and Actual Education Choices among Junior High School Students in Rural China: Evidence from Forty-one Impoverished Western Counties." *Chinese Education & Society* 46(4):30–42.
- Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology* 35(1):4–28.
- Steele, C., J. Aronson, and S. J. Spencer. 2007. "Stereotype Threat." *Encyclopedia of Social Psychology* 67:943–44.
- Steele, C. M. 2010. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. New York: W.W. Norton.
- Steele, C. M., and J. Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69(5):797–811.
- Steenbergen-Hu, S., M. C. Makel, and P. Olszewski-Kubilius. 2016. "What One Hundred Years of Research Says about the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-order Meta-analyses." *Review of Educational Research* 86(4):849–99.
- Stoet, Gijsbert, and David C. Geary. 2012. "Can Stereotype Threat Explain the Gender Gap in Mathematics Performance and Achievement?" *Review of General Psychology* 16(1):93–102.
- Stricker, Lawrence J., and William C. Ward. 2004. "Stereotype Threat, Inquiring about Test Takers' Ethnicity and Gender, and Standardized Test Performance." *Journal of Applied Social Psychology* 34(4):665–93.
- Van Bavel, Jay J., Peter Mende-Siedlecki, William J. Brady, and Diego A. Reinero. 2016. "Contextual Sensitivity in Scientific Reproducibility." *Proceedings of the National Academy of Sciences* 113(23):6454–59.
- Van de Werfhorst, Herman G., and Jonathan J. B. Mijs. 2010. "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective." *Annual Review of Sociology* 36:407–28.
- Walton, Gregory M., and Geoffrey L. Cohen. 2003. "Stereotype Lift." *Journal of Experimental Social Psychology* 39(5):456–67.
- Walton, Gregory M., and Geoffrey L. Cohen. 2007. "A Question of Belonging: Race, Social fit, and Achievement." *Journal of Personality and Social Psychology* 92(1):82–96.
- Walton, Gregory M., and Steven J. Spencer. 2009. "Latent Ability: Grades and Test Scores Systematically Underestimate the Intellectual Ability of Negatively Stereotyped Students." *Psychological Science* 20(9):1132–39.
- Wax, Amy. 2009. "Stereotype Threat: A Case of Overclaim Syndrome." In *The Science on Women and Science*, edited by C. Hoff-Sommers. Washington, DC: American Enterprise Institute.
- Wei, Thomas E. 2012. "Sticks, Stones, Words, and Broken Bones: New Field and Lab Evidence on Stereotype Threat." *Educational Evaluation and Policy Analysis* 34(4):465–88.
- Woronov, Terry. 2015. *Class Work: Vocational Schools and China's Urban Youth*. Stanford, CA: Stanford University Press.
- Yi, Hongmei, Linxiu Zhang, Chengfang Liu, James Chu, Prashant Loyalka, May Maani, and Jianguo Wei. 2013. "How Are Secondary Vocational Schools in China Measuring Up to Government Benchmarks?" *China and World Economy* 21(3):98–120.
- Zigerell, L. J. 2017. "Potential Publication Bias in the Stereotype Threat Literature: Comment on Nguyen and Ryan (2008)." *Journal of Applied Psychology* 102(8):1159–68.

### Author Biographies

**James Chu** is a PhD candidate in the Department of Sociology at Stanford University. His research focuses on the construction of new status distinctions and how the pursuit of status drives patterns of discrimination, aggression, and overwork in educational settings. James managed randomized controlled trials prior to graduate school.

**Prashant Loyalka** is an assistant professor (research) in the Graduate School of Education and a Center Research Fellow at the Freeman Spogli Institute for International Studies at Stanford University. His research focuses on examining and addressing inequalities in the education of youth and on understanding and improving the quality of education received by youth in large developing economies, including China, Russia and India.

**Guirong Li** is a professor in the School of Education of Henan University, Kaifeng, Henan Province, China. Her research interests are educational economics theory, efficient allocation of education resources, evaluation of teacher, and student development.

**Liya Gao** is a postgraduate student at Henan University, Kaifeng, Henan Province, China. Her major is educational economics and management.

**Yao Song** is an associate professor in the School of Education of Henan University, Kaifeng, Henan Province, China. His research interest is in the evaluation of educational policy.