Entrepreneurial Reluctance: Talent and Firm Creation in China*

Chong-en Bai† Ruixue Jia‡ Hongbin Li§ Xin Wang¶

Date: February 1, 2021

Abstract

We study whether talented individuals are more or less likely to create firms, linking administrative college admissions data for 1.8 million individuals with information on firm registration records in China. The data reveals that most of the variation in firm creation is from within college. Given the same college, we find that individuals with higher college entrance exam scores – the most important measure of talent in this context – are *less likely* to create firms. Moreover, this negative relationship is even stronger for groups with social and economic advantages such as men, urban individuals, and those from better high schools. We propose three hypotheses for why students with higher-scores are less likely to create firms: the lower entrepreneurial ability hypothesis, the opportunity cost hypothesis and the risk attitude hypothesis. The data favors the opportunity cost hypothesis (i.e. higher-score individuals are attracted away by non-entrepreneur sectors, even though they might have *higher entrepreneurial ability*). Moreover, the exam score to firm creation relationship varies greatly across province and industry according to the importance of the state sector, indicating the scope for improvement in talent allocation.

---

*We thank Ying Bai, Ting Chen, Julie Cullen, Gordon Dahl, Roger Gordon, Yona Rubinstein, Imran Rasul, Eric Wang, Xiabo Zhang, and participants of seminars at Fudan, Georgetown-World Bank, LSE, NBER, UBC, UCL, UCSD, USC for their comments.

†Tsinghua University, baichn@sem.tsinghua.edu.cn

‡LSE and UCSD, rxjia@ucsd.edu.

§Stanford University, hongbinli@stanford.edu

¶CUHK, wangx2.04@gmail.com

1 Introduction

In recent years, a growing literature in economics has shown that entrepreneurial talent has important influences on both firm-level productivity and aggregate-level economic growth (e.g., Bertrand and Schoar 2003; Bloom and Van Reenen 2007; Syverson 2011; Gennaioli et al. 2013), suggesting that it is important to allocate talents to the entrepreneurial sector. However, it remains unclear whether talented individuals in a society are more or less likely to become entrepreneurs. Some argue that talent facilitates entrepreneurship especially in terms of incorporated business (as opposed to self-employment, e.g., Levine and Rubinstein 2017, 2019), while others emphasize that talents are often attracted to non-entrepreneurial sectors in many societies which can lead to misallocation of talent (e.g., Murphy, Shleifer and Vishny 1991; Baumol, Litan and Schramm 2007). Empirically examining this link is challenging. One would need a dataset that includes good information on both talent and entrepreneurship, in addition to large data samples given the small probability of firm creation. Most data does not even have enough information for researchers to define entrepreneurship or firm creation because it cannot differentiate big firms from small businesses.[1] Moreover, it is also generally difficult to be able to separate the impact of ability from that of education and family background.

In this paper, we examine whether and how talent affects entrepreneurship or firm creation using administrative data from China.[2] Empirically, we link college admission records in 1999-2003 with the universe of Chinese firms and their owners, and use a random sample of 20% of the linked administrative data (over 1.8 million college graduates who had created around 170,000 firms by 2015) to examine who had created firms by 2015 and their firms' respective performance. We can then define entrepreneurship more precisely and paint a comprehensive picture of the creation and performance for both small and large firms by observing firm size and other firm performance measures.

We focus on a specific measure of talent, one's score in the National College Entrance Exam (known as *gaokao*), as the criterion in the centralized college admission system – popularly believed to determine the course of one's life. Admittedly, it is an imperfect metric, as it mainly measures

[1] A recent literature points out the importance of differing big firms from small business as only the former likely to involve transformational entrepreneurship (e.g., Schoar 2010).

[2] Entrepreneurship is a multiple discipline topic. We do not give an overview of the extensive literature here. See Parker (2018) for many related studies in economics, finance, and other fields.

a student's cognitive skills. However, it is a measure in which many societies (e.g., China, Japan, South Korea, Turkey) have already invested heavily, there remain contentious debate about what such scores capture. Thus, it seems important to understand whether this measure capture a general ability that is positively associated with entrepreneur ability. Our study speaks to this debate.

We focus on within-college analyses, for both conceptual and empirical reasons. Conceptually, we leverage our data to separate the role of talent from that of education. The within-college design allows us to remove the effects of college reputation and network. Empirically, we find that most of the variation in firm creation comes from within colleges. For instance, while the collect fixed effect can explain up to 19% of the variation in wages of paid jobs (in our survey data), they can only account for 1.2% of the variation in firm creation. In addition, we find that the college fixed effects can only explain 46% of the variation in exam scores, leaving the majority of the variation to occur within colleges.[3]

We find that higher-score students are less likely to become entrepreneurs. First of all, the raw correlation between firm creation and the *gaokao* score is negative. Most importantly, after removing the college fixed effects, we still find a strong negative association of firm creation and score within a college: a one standard deviation higher score is associated with a 10% lower probability of creating a firm. Note that this association has nothing to do with time/effort allocation in college as the score is determined prior to entering college. Within colleges, this negative sore-firm creation relationship holds even when we focus on the creations of more successful (larger) firms, compare students within college-field, and examine colleges by tier. We also find that firm creation is positively correlated with observable family background variables (which are in turn positively correlated with scores), suggesting that unobserved family background variables, if any, would not have been underlying the negative correlation between the score and firm creation. Altogether, these results suggest that talented students are less likely to become entrepreneurs, holding their college education and family background constant.

We further find that the negative score-firm link is stronger for men, urban students and those from better-quality high schools. This means that the negative link appears even stronger for the relatively more advantaged groups, because we also find that students who are male, urban and from better high schools are more likely to create firms than their counterparts . The slope difference is economically large across gender and high-school quality. For instance, the negative score-firm creation slope for men (and those from above-median high schools) is 67% (and 50%) higher than women (and those from below-median high schools).

[3]This is probably due to the uncertainty in the admission system and political economy of provincial quota assignment.

We then attempt to explain the negative link between exam scores and firm creation by formalizing three hypotheses in a simple conceptual framework built on earlier work (e.g., Evans and Jovanovic 1989; Levine and Rubinstein 2019). Specific to our context, we allow exam scores to map to the entrepreneurial ability and wage-job ability differently. The first hypothesis is the *"lower entrepreneurial ability"* hypothesis, which asserts a negative correlation between exam scores and entrepreneurial ability. For instance, one can assume that those with higher scores are the opposite type of Jack-of-all-trades (e.g., Lazear 2004) or that the exam-focused Chinese education system deters entrepreneurship. The second is the *"opportunity cost"* hypothesis. In this case, scores are positively correlated with both entrepreneurial and wage-job abilities, and the talent gets attracted away by the waged jobs. The final hypothesis is the *"risk attitude"* hypothesis, which argues that higher-score students are more risk-averse than their lower-score peers in a college. To test these hypotheses, we supplement the administrative data with our own survey of Chinese college graduates we conducted during 2010-2015, covering detailed information on their first job outcomes.

Our data does not support the negative entrepreneurial ability hypothesis. To test this, we examine firm success and find that firms created by higher-score students are larger, more likely to expand and even more likely to become publicly listed. These patterns are robust to using peer composition – whether one happened to have more peers from more entrepreneurial provinces in the same cohort within a college – as an instrument to correct potential selection bias of firm creation. A survival analysis further shows that the hazard ratio of exit is lower for firms founded by higher-score students. All the patterns reveal that a higher score indicates higher entrepreneurial ability rather than the opposite.

The empirical results fit the opportunity cost hypothesis best. Using our survey data, we document a positive link between scores and first-job wages within colleges, suggesting the opportunity cost of starting up own businesses increases with scores. Moreover, we find that the positive score-wage link is also stronger for men, urban students, and those from better high schools, which mirrors the pattern of firm creation in the administrative data. Altogether, these results suggest that the steeper negative slope of score-entrepreneurship relationship for students with advantageous social economic backgrounds is because students in this group are more likely to be lured away by opportunities of paid jobs. When we further separate waged jobs in the state sector (state-owned enterprise and government jobs) and the private sector, we find that both are relevant as alternative opportunities but the state sector plays a relatively larger role in attracting the talents.

We do not find evidence for the risk aversion hypothesis. Empirical patterns on risk attitudes in our supplementary survey data confirm the intuition that women are more risk averse than men,

suggesting that our risk measure is informative. In contrast, they reveal no strong correlation between risk aversion and scores in a college. In addition, there are no clear heterogenous score-risk patterns by personal characteristics that can explain the heterogenous patterns on score and firm creation by personal background. To be clear, this finding is not to belittle the important role of risk preferences in firm creation but to show that it is difficult to use risk preferences to explain our findings on score-firm relationships.

Overall, our findings imply that talent measured by entrance exam scores, the most important human capital measure in the Chinese education system, is positively associated with both entrepreneurial ability and wage ability. Many societies rely on exams to regulate access to education, which shapes how resources are allocated to the education system. Yet, little is known about the relationship between exam scores and ability. Our paper is one of the first to shed light on this issue.

When it comes to talent allocation across occupations, we find more evidence for the argument that higher-score students are more likely to be attracted away by non-entrepreneurial sectors than their lower-score college peers, which provides evidence for earlier theoretical conjectures that talent was more likely to be allocated to rent-seeking sectors rather than the entrepreneur sector (e.g., Murphy, Shleifer and Vishny 1991; Baumol, Litan and Schramm 2007).

Finally, we document that the relationship between score and firm creation varies greatly across province and industry. For instance, a one standard deviation higher score is associated with a 19.6% lower probability of creating a firm in Xinjiang but 6.7% in Zhejiang. Analogous to the insight in Hsieh and Klenow (2009), the wide variation suggests the existence of misallocation. While many factors are underlying these variation, state penetration (measured by the share of state investment in total fixed investment) appears to be systematically associated with more entrepreneurial reluctance of higher-score individuals. These results indicate a large scope for improvement in talent allocation: more higher-score individuals will enter the entrepreneur sector with a retreat of state intervention in the economy. These findings join the recent literature using data to understand talent misallocation in different settings (e.g., Hsieh et al. 2019 on race discrimination in the U.S. and Ashraf et al. 2021 on gender norms across countries). Ours highlights that the state plays an important role in shaping talent allocation.

2 Background and Data

2.1 *Gaokao* and Firm Owner Data

All 2,000 or so colleges in China admit students using a centralized system that revolves around the annual college entrance exam (i.e., *gaokao*), taken in early June every year. The exam, written and graded by provincial education authorities, determines the college, if any, a student will attend. In the first or second year in high school, students are divided into natural sciences (with exam subjects being Chinese, English, math and sciences) and social sciences (with exam subjects being Chinese, English, advanced math and social sciences) tracks. After the third year, they take the exam in one track. Thus, these exam scores are comparable only for students from the same province-year-track (social or natural sciences).

We employ administrative data covering the whole universe of all participants in *gaokao* from 1999-2003 and their college admission outcomes across 2,056 colleges. Out of the 23 million records, including those who failed to go to any college, 14.2 million were accepted by some college, which is the data we use. The data provides detailed information on students' exam performance including total scores, subject scores, names of colleges and majors (if admitted), and students' biographical information such as gender, *Hukou* (i.e., rural or urban residency), birth year and high school names.[4]

To track firm creation and performance, we use administrative data of all firms ever registered in China by February 2015.[5] This data includes 28 million firms, 11 million of which had been de-registered by 2015. The administrative registration records provide information on the owners (known as shareholders), which can be individuals and firms. We call individual owners firm creators or entrepreneurs. We also know if a firm invested in another firm as an owner, which we define as an "expansion". The data also includes some basic firm information such as industry, firm location, and registered capital size. In addition, we observe when a firm is deregistered, which we then use as a proxy for exit.

[4]It should be noted that the college dropout rate is very low in China and thus is not an important issue in our context. In fact, the college education system in China is known for being "strict entrance, easy out". The overall graduation rate for Chinese colleges is above 95%.The Beijing-based Mycos Institute estimated in 2011 that China's college dropout rate is 3%, while the Ministry of Education said that year that the rate is 0.75%.

[5]For small business, an individual may face a trade-off between staying self-employed (not studied in this paper) and registering a firm. Putting aside the fact that certain industries require registration for doing business, an important benefit of registering a firm is limited liability. On the other hand, staying self-employed without registering a firm may have implications that contribute to tax evasion. We consider both small and big firms, registration is a must for the latter.

The college entrance exam and admission data are linked to the firm owner data by an encrypted national identity number in both datasets. Out of the 14.2 million records, we succeed in linking 9.5 million records,[6] with 9.1 unique individuals. The extra 400,000 records are repeated exam takers, for whom we focus on the first-time score because we are interested in score as a proxy for ability.[7] For the purpose of de-identification, we randomly sampled 20% of the college students (i.e., 1.8 million individuals) in the linked administrative data for our analyses.

2.2 College Graduates Survey 2010-15

We supplement the administrative data with our own college graduate survey data, which have information on wages. We conducted large scale surveys of college graduates during the graduation months (May and June) in 2010-15, which cover around 30,000 students from 90 colleges, 14,600 of whom reported rich information on their first-jobs. We designed the surveys to evaluate elite college premium and intentionally asked about their *gaokao* scores (see Jia and Li (2019) for a detailed description of these surveys and where we also show that the first job has a strong persistent effect on future jobs). Moreover, the surveys include rich information on student performance and behaviors in college. Although these surveys only cover a limited number of colleges, this is not a major issue for us since our main interest is comparison of students who attend the same colleges.

2.3 Variables and Summary Statistics

Firm Variables We present firm-level variables in Panels B and C of Table 1. By 2015, the 1.8 million college graduates in our data had established 170,087 firms, and the probability of creating any firm is 7.2%. The median firm was established in 2010, or around six years after college graduation. The top five industries are wholesale and retail (30.9%), leasing and business services (20.2%), scientific research and technology services (14%), manufacturing (8.5%), and information technology and services (6.8%). The remaining 15 industries accounted for 20%.

In light of recent literature emphasizing the importance of large firms with transformative entrepreneurship, it is important to consider firm size. We use registered capital as a proxy for firm size,[8] which is the maximum liability a firm has and hence matters for doing business. After a firm is registered, the owners can choose to update their registered capital size, but we can only observe

[6] Matching of the two data is not perfect because of missing (3.9 million) or invalid identifying numbers (0.8 million).

[7] We examine how missing data due to linking and repeated exam takers affect our results. In addition, a small share of students got some extra points due to none-exam experience, which we exclude from our score measure.

[8] Unfortunately, the firm registration data does not report employment.

the most current number in 2015. If a firm exited before 2015, its registered capital size refers to the most recent information before exit. Using registered capital as a measure of firm size, we find that 25th and 75th percentiles are RMB 200,000 and 2 million, about 1.6% of the college graduates had established firms over RMB 2 million, and only 0.4% above RMB 15 million.

Besides firm size, we also use a few alternative variables to measure firm success. The first is whether the firm was registered out of one's home province. It has been well recognized that provincial governments tend to protect local firms from the competition of other firms (e.g., Young 2000). Thus, being able to start a firm out of one's home province indicates success. In our sample, less than half (or 39%) of the firms were registered outside one's home province, suggesting that it is not easy to establish a firm beyond one's home province due to local protection. The second is expansion, which is defined as one firm investing in another firm as one of owners. n our sample, 5% of firms expanded this way. The third proxy is whether a firm became publicly listed. This measure can be considered as a proxy for extreme success as it is a rare event with a chance of 3.6 per 10,000 firms. Finally, we use deregistration information to proxy firm exits. The exit probability varies greatly by firm age: 90% survived beyond 3 years but only 60% survived beyond 10 years.

It is interesting to note that most of the variation in firm creation comes from within colleges. College fixed effects only account for 1.2% of the variation in firm creation. To benchmark this, we find that college fixed effects explain over 19% of the variation in wages in our college graduate survey.

Exam Scores As expected, exam scores are positively correlated with college quality. Figure 1(a) shows the distribution of exam scores (after controlling for province-year-track fixed effects) of eight college groups (top 10, 11-20, 21-50...). On average, better colleges accepted higher-score students. However, college fixed effects can only explain less than half (46%) of the variation in exam scores. Exam scores actually vary greatly within colleges, evidence by the overlapping of scores across different batches of colleges in Figure 1(a).

To measure the variation, we report in Panel A of Table 1 the mean and standard deviation of the exam scores, which are 443.8 and 101.5. Once we control for province-year-tracks fixed effects, the standard deviation drops to 87.7. It goes down further to 65 after we also control for the college fixed effects, and 63.1 if we add 12 major fixed effects. Similarly, Figure 1(b) also shows a remaining large variation of the score after we control for all the fixed effects.

The large within-college variation of scores might be due to a few institutional reasons. First, college application and admission are a very uncertain process. In our study period, students in

most provinces applied for colleges before they knew their exam scores. Each exam taker needed to fill a college preference sheet by pencil and bubble sheet for up to three colleges (and three majors in each college) within a few days. Each student could be accepted by only one college, and priority was given to the first choice in the bubble sheet (second and third choices were almost useless). As a result, the matching between score and college was far from being ideal.[9] Second, scores for each college varies greatly across provinces due to the uneven distribution of admission quota. Each province is assigned a quota for each college by the central government. Due to political and historical considerations, major metropolitans such as Beijing, Shanghai, and minority provinces such as Tibet, Xinjiang and Yunnan typically get a larger quota, especially for elite colleges. Finally, there is also a college-major tradeoff. Some students may choose a lower-ranked college for a popular major. For example, within any college, admission scores for popular majors such as economics, finance, law and STEM are normally higher than those in humanity. We will consider majors in our analyses.

Personal Background We report summary statistics of personal attributes in Panel A of Table 1 and their correlations with *gaokao* scores in Table A. The median birth year for these cohorts is 1982, i.e., the median age is 33 in 2015 (the final year of the firm registration records we examine). As shown by Table 1, 54% of the colleges students are male, slightly higher than the male share in the population (51.3% in 2001). 53% of the student have urban *Hukou*, much higher than the urban share in the population (37% in 2001), consistent with the fact that fewer rural students can make it to high schools. As shown in Table A, within college, females and males do not differ in the *gaokao* score, though men do better across colleges. Rural students have higher scores, both across and within colleges, most likely a result of only the better rural students being offered the opportunity to attend high schools and participate in the exam.

Because we do not observe family socioeconomic status, we use high school quality as a proxy. Richer families are more likely to afford more elite high schools whose quality is determined at the local level. For instance, Ye (2015) shows that the parents of those in better high schools are better educated and have better-paying jobs. We measure high school quality using the share of students admitted by the top 100 colleges in China[10] for each high school. As reported in Table 1, the bottom 25% and top 25% high schools have 4.8% and 19.4% of their students admitted by the top 100 colleges, indicating wide variations across high schools. In addition, we use one's birth

[9]Such uncertainty gets marginally mitigated only in recent years when the admission system was reformed to be a parallel system which allows students to apply for a few more colleges thanks to computer technologies.

[10]These are designated by the government as the Project-211 colleges – the top 100 colleges in the 21^{st} century.

county's GDP per capita in 2001 as another indirect measure of one's family economic status. Not surprisingly, students from better high schools and richer counties tend to have higher scores, both across and within colleges, as shown in Table A.

3 Talent and Firm Creation

In this section, our focus is to understand the empirical relationship between individual-level ability and firm creation. Is it positive or negative? How large is it? Does it vary by personal background?

3.1 Descriptive Results

We first visualize the relationship between college entrance exam score and firm creation both unconditional and conditional on the colleges, recognizing that the former relationship reflects both the within- and between-college effects. We have three different definitions of firms based on the size of capital, i.e., any firm, a firm with a registered capital no less than RMB 2 million or 15 million. We isolate province-year-track fixed effects so that the scores are comparable.

Descriptive results suggest that the correlation between firm creation and exam scores is generally negative, but even more so when we control for college fixed effects. As shown in Figure 2(a), the score-firm correlation is strongly negative when all firms count. This negative correlation is even larger within colleges (Figure 2(b)). Moreover, within colleges, regardless of how we define firms, higher-score students are less likely to create them, as evidenced by Figures 2(c) and (d).

3.2 Within-College Estimation Results

Baseline To estimate the link between academic talent and firm creation within colleges, we employ the following specification at the individual level:

$$Firm_{i,pyt,c} = \beta Score_{i,pyt,c} + \alpha X_i + \lambda_{pyt} + \theta_c + \epsilon_{i,pyt,c}, \tag{1}$$

where $Firm_{i,pyt,c}$ a dummy variable indicates whether individual i of province-year-track (pyt) in college (c) created a firm. On the left side of the equation, we use four different definitions of firms by size. The key independent variable of our interest is the exam score ($Score_{i,pyt,c}$). X_i indicates one's personal characteristics that include gender, *Hukou* (rural vs. urban), high school quality, birth county's GDP per capita, and age fixed effects. While we do not have measures like parental income, it is reasonable to assume that those from better high schools and richer counties have higher socioeconomic status. Since $Score_{i,pyt,c}$ is only comparable within province-year-track, we

always control for province-year-track fixed effects (λ_{pyt}) in our analysis. θ_c indicates college fixed effects. We report standard errors are clustered at the college level. We also control for the major fixed effects in X_i for certain specifications.

Regression results reported in Table 2 show that firm creation is negatively correlated with exam scores. To make interpretations easier, we report the coefficient for 1σ (standard deviation) difference in the exam score (i.e., 65 points in a college). As shown by Column (1), within a college, 1σ higher exam score decreases the probability of creating any firm by 0.77 percentage points or over 10% of the mean (7.12). The estimates remain stable when we control for all personal characteristics in column (2) and add 12-major fixed effects in column (3). If we define firms as those with registered capital no less than RMB 2 million or 15 million, the magnitudes of the estimates are actually very similar as reported in columns (4)-(9).

In addition, we find personal attributes matter for firm creation. The gender difference is large: women are only half as likely to create a firm as men, probably because female students are more risk averse or face discrimination in access to capital. Students from urban areas and better high schools are more likely to create firms than their counterparts, albeit the difference is less striking than the gender difference, suggesting that students with advantageous social and economic background also have an advantage of creating firms.

Potential Data Issues We check how sensitive our analysis is to a number of data issues. First, to check whether missing data due to matching is an issue, we restrict our analysis to the province-years with missing probabilities lower than 5% and 1%. As reported in panels A and B of Appendix Table B.2, results remain similar, suggesting that the missing identification may not have been a critical issue in our case. Second, we exclude repeated exam takers from our analysis, to examine whether including them has caused a problem. Again, we obtain estimates very similar to those from our baseline estimations (reported in Appendix Table B.3).

Third, as indicated in Figure 1, there are a small number of students on the tails of the within-college score distribution. Although, the negative link through the whole range of scores in Figure 2(b) suggests outliers may not be an issue, as another test, we carry out our analysis removing 10% of observations from both tails. Estimated negative correlations become even stronger, as shown in Appendix Table B.4.

3.3 Heterogeneity Across Groups

By Major and College Rank Within a college, the opportunity cost of creating a firm may differ across other attributes such as major. We controlled for major dummies in the above analysis and find firm creation-score correlation holds. We now go a step further to allow the effect of scores on firm creation to vary across major. To track our analysis, we categorize the twelve majors into three groups: STEM, Economics-Finance-Law, and Humanity. STEM majors account for 51% of college students, consistent with the fact that Chinese higher education encourages a large number of students to enter STEM fields. Economics, Finance and Law majors account for 27% of the students and humanity majors for the rest 21%.

The relationship between exam scores and firm creation within each broad major is similar to the overall effect reported above (Table B.5). First of all, as shown by the mean firm creation probabilities, students in Econ-Finance-Law majors have a higher propensity of firm creation, consistent with the conjecture that the choice of these majors indicates preferences toward entrepreneurship. Most importantly, the negative relationship between scores and firm creation holds within each major. For instance, based on the broadest definition of a firm, 1σ higher exam score within a college decreases the probabilities of creating firms by 10% for those who majored in STEM, 7.6% for those in Economics-Finance-Law majors, 12% for those in Humanity.

The negative correlation is also not driven by colleges of certain rank. Theoretically, the relationship between scores and firm creation may differ by college quality, if the opportunity cost of firm creation differs greatly across colleges of different quality. To test this, we divide colleges into three groups: the top 10, those ranked 11-100, and the rest. As reported in Appendix Table B.6, the negative link between scores and firm creation hold for each group.

By Social and Economic Variables We further examine whether and how the link between scores and firm creation varies by personal characteristics, which may help us to form hypotheses explaining the negative correlation between firm creation and scores. Specifically, we examine the score-firm creation link across gender, *Hukou* (rural vs. urban), and high school quality using the following specification,

$$Firm_{i,pyt,c} = \beta_1 Score_{i,pyt,c} + \beta_2 Score_{i,pyt,c} \times \mathrm{X}_i + \lambda_{pyt} + \theta_c \times \mathrm{X}_i + \epsilon_{i,pyt,c}, \tag{2}$$

where X_i indicates being female, rural or from below-median quality high schools. We also allow college effects to differ for these groups by including $\times \mathrm{X}_i$ as controls.

Results seem to suggest that the negative link between firm creation and score is stronger for

individuals with more advantageous background. As shown by Figures 3(a)-(c), the within-college negative correlations between scores and firm creation is stronger for men, urban individuals, and for those from better high schools.[11]

The differences of the estimated correlation between firm creation and score are large across personal background groups. For instance, the estimates in column (1) of Table 3 show that the score-firm slope for men is 65% higher than that for women.[12] The negative score-firm link is also stronger for urban than rural, though the difference is less striking (column (2)). The difference across high schools is also large: the magnitude for above-median high schools is about 50% larger than that for below-median high schools (column (3)). Results do not change much when we include all these variables together (column (4)) and restrict firms to be the larger ones (columns (5)-(12)).

3.4 Discussion

There are three potential hypotheses that could explain our finding that higher-score students in a college are less likely to create a firm. The first is that students with higher scores in college have lower entrepreneurial ability than their lower-score counterparts, which we call the *"lower entrepreneurial ability"* hypothesis. While we are not aware of any evidence for such an argument, the perception is that the Chinese education system hinders creativity and entrepreneurship.[13] The second hypothesis is the opposite: higher-score students in college have higher entrepreneurial ability but they also have higher wage ability. If the return of the wage ability to score rises faster than that of the entrepreneurial ability, we expect a negative relationship between scores and firm creation. We term this the *"opportunity cost"* hypothesis. The third hypothesis is that scores reflect other personal traits of entrepreneurship, and in particular, risk attitude is an important one. One may wonder whether higher-score students in a college are more risk averse than their lower-score peers. We call this the *"risk attitude"* hypothesis. We formalize all three hypotheses in Section 4 and use our data to test which is/are relevant in our context.

[11] For simplicity, we focus on firms with registered capital no less than RMB 15 million in these figures and report similar patterns by firm size in Table 3.

[12] 1σ higher score within a college decreases the probability of creating any firm by 0.96 for male but 0.58 for female.

[13] There exists a Chinese phrase to describe this perception, *gaofen dineng*, meaning "high score but lower ability". Zhao (2009) identifies this as a problem with the Chinese educational system. However, we are not aware of any empirical investigation to prove the plausibility of this perception.

4 A Conceptual Framework

In this section, we present a simple model to derive the three testable hypotheses on the link between firm creation and academic ability. Building on earlier work (e.g., Evans and Jovanovic 1989; Levine and Rubinstein 2019), the most important input is the academic ability, which feeds into entrepreneurial ability and wage ability (ability for wage work) differently. Specific in our context, the exam score is the indicator of academic ability.

Setup An individual with score S (we ignore the subscript i to simplify notations) and personal background X (e.g., gender and *Hukou*) chooses to either establish a firm or become a worker. Score further feeds into two different abilities for work, the entrepreneurial ability (A_E) and wage ability (A_W). We allow for a flexible mapping between the score and these abilities, specified as

$$A_E = e^{aS} \text{ and } A_W = e^{bS}. \tag{3}$$

a can have any sign, reflecting the correlations between the exam score and entrepreneurial ability can be positive, negative or zero. In theory, we could allow b to have any sign, but for empirical relevance, we assume $b > 0$ (on which we later present evidence).

If one chooses to start a firm, then the entrepreneurial ability (A_E) and physical capital K are used to produce the output Y with the production function

$$Y = A_E^{\beta(X)} K^{\alpha}.$$

β captures the return to entrepreneurial ability, which is allowed to be affected by personal background (X) in a flexible way.

A firm owner maximizes the following expected returns by choosing the optimal size of capital K,

$$R = A_E^{\beta(X)} K^{\alpha} - r(X)K,$$

where the price of capital $r(X)$ can also depend on personal background. Solving this, we get the optimal amount of capital

$$K^* = \left(\frac{\alpha A_E^{\beta(X)}}{r(X)} \right)^{\frac{1}{1-\alpha}}. \tag{4}$$

Given that $A_E = e^{aS}$, we obtain the expected returns as

$$\begin{aligned} \ln R^* &= \tfrac{1}{1-\alpha} \ln A_E \beta(X) - \tfrac{\alpha}{1-\alpha} \ln r(X) + \ln \Gamma \\ &= \tfrac{1}{1-\alpha} aS\beta(X) - \tfrac{\alpha}{1-\alpha} \ln r(X) + \ln \Gamma, \end{aligned} \tag{5}$$

where $\Gamma = \alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}$.

Alternatively, if one chooses to become a wage worker, then wage is a function of wage ability A_W or

$$W = A_W^{\mu(X)},$$

where μ captures the return to wage ability, which is also allowed to be affected by personal background (X) in a flexible way.

Similar to the existing literature, we capture personal preferences such as risk aversion in occupation choice in a reduced-form way. Specifically, we assume the utility function includes an element $\delta(X, S)$, where δ refers to the extra utility of being a waged worker (or utility of avoiding risks of firm creation). Thus, the utility of being a worker can be written as

$$V = A_W^{\mu(X)} e^{\delta(X,S)},$$

and taking a log we get

$$\begin{aligned} \ln V &= \mu(X) \ln A_W + \delta(X, S) \\ &= bS\mu(X) + \delta(X, S). \end{aligned} \tag{6}$$

Choosing to Create a Firm An individual will choose to create a firm iff $lnR^* \geq lnV + \varepsilon$, where we also allow individual-level idiosyncratic shocks (ε) to affect occupation choice and ε follows a distribution of c.d.f $\mathbb{G}(\varepsilon)$. We then derives

$$\frac{1}{1-\alpha} aS\beta(X) - \frac{\alpha}{1-\alpha} \ln r(X) + \ln \Gamma \geq bS\mu(X_i) + \delta(X, S) + \varepsilon. \tag{7}$$

Thus, the probability of creating a firm becomes

$$\mathbb{P}(Firm) = \mathbb{G}\left\{ \left[\frac{1}{1-\alpha} a\beta(X) - b\mu(X) \right] S - \delta(X, S) - \frac{\alpha}{1-\alpha} \ln r(X) + \ln \Gamma \right\}. \tag{8}$$

From equation (8), we can see a negative link between scores and firm creation (dP/dS) is equivalent to the following inequality,

$$\frac{1}{1-\alpha} a\beta(X) - b\mu(X) - \delta_s(X, S) < 0. \tag{9}$$

The negative link between scores and firm creation ($d\mathbb{P}/dS$) could be due to three channels:

(1) the *"lower entrepreneurial ability"* hypothesis or $a < 0$;

(2) the *"opportunity cost"* hypothesis refers to $b > 0$;

(3) the *"risk attitudes"* hypothesis refers to $\delta_s(X, S) > 0$.

Testable Implications We can derive a few empirical tests of the three hypotheses based on (4) and (9). Some or all of the three hypotheses could be at work to drive the negative correlation between creating a firm and score S. In the following, we derive a few testable conditions that make

(9) more likely to hold. Note that individually, these tests results are neither necessary nor sufficient conditions for (9) to be true.

First, we can test the *"lower entrepreneurial ability"* hypothesis by estimating the sign of a. Empirically, we do so by estimating the effect of score on the optimal size of capital, which is derived from Equation (4),

$$\ln K^* = \frac{1}{1-\alpha}\left[aS\beta(X) + \ln\alpha - \ln r\right]. \tag{10}$$

The estimated sign of dK/dS is the same as a, given that $\beta > 0$. If the estimated sign is negative (score is negatively correlated with the size of capital), then $a < 0$, which means that we cannot reject the *"lower entrepreneurial ability"* hypothesis.

Second, we can test the *"opportunity cost"* hypothesis by estimating the sign of b, the wage ability. Particularly, we estimate $lnA_w = bS$, or log wage as a function of score S, exploring our college graduate survey. Finding a positive correlation ($b > 0$) would provide supports of the *"opportunity cost"* hypothesis.

Third, we can test the *"risk attitude"* hypothesis by estimating the sign of $\delta_s(X, S)$, the correlation between risk attitude and score, using survey data with both score and risk attitude information.

Fourth, we can also go a step further to test how personal characteristics (X) affect the marginal effect $d\mathbb{P}/dS$, or estimating the sign of the cross derivative $\frac{d^2\mathbb{P}}{dSdX}$. For example, if we find that the score-risk relationship does not change with X but the score-wage relationship changes with X, then it supports the *"opportunity cost"* hypothesis.

5 Testing the Hypotheses

In this section, we test the three hypotheses derived from the model. In theory, all three hypotheses could work together.

5.1 Lower Entrepreneurial Ability Hypothesis

Estimation Strategy We test the *lower entrepreneurial ability* hypothesis by recovering the sign of a, which is the same as dK/dS. Empirically, we estimate equation (10) using the following specification,

$$\ln K_{f,i,pyt,c} = \gamma Score_{f,i,pyt,c} + \alpha X_i + \lambda_{pyt} + \theta_c + \epsilon_{i,pyt,c}, \tag{11}$$

where $K_{f,i,pyt,c}$ refers to registered capital of firm f created by individual i.

One concern is selection bias. Success is conditional on entry, which may be correlated with unobservable variables that can also affect success. For instance, let us assume that individual-level shocks in the decision to create a firm (inequality (7)) are not correlated with academic talent in our conceptual framework but we cannot empirically observe these shocks. If they vary systematically with academic ability, the estimation of equation (11) is biased.

We employ the Heckman two-stage estimation to correct potential bias caused by entry (Heckman 1976). In the first stage, we estimate the probability of entry, which is

$$Pr(Firm = 1|Z_{i,pyt,c}) = \Phi(Z_{i,pyt,c}), \tag{12}$$

where $Z_{i,pyt,c}$ includes X_i, λ_{pyt}, θ_c in equation (11) as well as an instrument variable (IV). We then estimate the second-stage firm success equation correcting for (entry) selection by using the inverse mills ratio estimated from the first stage.

We argue that the origin composition of peers within the same college-cohort (indicated by peer origin$_{c,t}$) provides a reasonable instrument to predict firm entry. Essentially, we exploit two sources of variation: the variation in entrepreneurship across provinces and the composition of one's peers within a college. Intuitively, within a college, the composition of students' origin has some randomness which creates the variation in entrepreneurship exposure.

Specifically, we define the instrument as

$$\text{peer-origin}_{c,t}=\left(\frac{\sum_{j\neq i} EntrepreneurProp_{j,prov}}{n-1}\right)_{,c,t}$$

where $EntrepreneurProp_{j,prov}$ is calculated as the number of entrepreneurs who had established firms before 1999 divided by the adult population in a province. We use this historical information of entrepreneurial propensity to further minimize any concern of the reflection problem. n refers to the number of students in a college-cohort.

A good IV should have a strong first stage results, i.e., they should be highly correlated with firm creation (selection). We indeed find that, within a college, peer-origin$_{c,t}$ is a statistically significant predictor of firm creation (column (3) of Table 4. In terms of the marginal effect, 1σ increase in $peer_{c,t}$ is associated with a 2.4 percentage points higher probability of firm creation, or 25% of the mean.

Results on Success Using capital sizes as success measures, we find evidence that success of an entrepreneur increases with scores. Employing log capital size demeaned by industry as the outcomes, we find positive associations between scores and registered capital (columns (4)-(5) of

Table 4). The estimated coefficients suggest that 1σ increase in within-college scores is associated with RMB 220,000 larger registered capital within an industry. These coefficients are relatively small (around 0.5-3% of the mean).

The positive links are stronger when we use alternative measures of firm success, including a firm being located outside one's home province, firm expansion, and becoming publicly listed. For these three alternative measures, we find positive and sizable correlations between exam scores and firm success (columns (6)-(8)): 1σ increase in the within-college exam score increases the probabilities of investing out of home province and expansion by around 10% (i.e., 0.039/0.394 and 0.005/0.049). When it comes to the rare event of becoming listed, the estimate is less precise but the increase is even larger in magnitude, around 17% (0.063/0.359) of the mean.

As a comparison, we also report simple OLS estimates (conditional on entry) without correcting for selection. Results reported in panel B of Table 4 and Appendix Figure C.2 show that the OLS estimates are similar to the Heckman estimates, suggesting that the selection bias is not quantitatively important, conditional on a valid IV.

Exclusion Restriction A good IV should be randomly assigned and thus not correlated with other covariates of success. This is indeed the case. In Table 4, we show that across colleges, score is positively correlated with peer-origin$_c$ (column (1)), indicating students sorting across colleges; however, within a college, the correlation between peer-origin$_{c,t}$ and score is close to zero (columns (2)). This is also true for other personal characteristics, as reported in Appendix Table C.1. The finding of no correlation between our instrument and observables (scores and other personal characteristics) gives us confidence that the IV is also unlikely to be highly correlated with unobservables.

As another exclusion restriction test, we create two IVs by the origin of peers, those from the same versus different regions (region means northern or southern China). If unobservables are important, then we should expect to see a larger two-step estimate when using peer exposure from the same region as the IV because these peers are more likely to be exposed to the same environments that affect success. Our results in Appendix Table C.3 show that the magnitudes of the estimates are very similar when using the two IVs, suggesting that the effect of unobservables are not quantitatively important.[14]

[14] Another similar exercise to divide peers by their success, assuming that less successful peers are less likely to help one's own firm. We find similar patterns using more or less successful peers to predict entry. We did not present these results because this criterion is based on ex post measure, which is less ideal than our current measure based on ex ante proxy, i.e., province of origin.

A Survival Analysis As an additional success (or failure) measure, we conduct a survival analysis to study how within-college scores correlate with firm exits (deregistration in our data). We plot the firm survival probabilities for owners with within-college scores in the top 20% in relation to those in the bottom 20%.[15] As shown in Appendix Figure C.4, firms founded by those with higher scores are more likely to survive, and the advantage rises over time. In a Cox regression, the estimated hazard ratio with respect to 1σ within-college exam scores is around 0.975, with a standard error of 0.006. The small-than-1 ratio is consistent with the graphical illustration that firms created by higher-score students are less likely to exit.

Summary The empirical evidence seems to reject the *lower entrepreneurial ability* hypothesis. In other words, we find students with higher scores create more successful firms, suggesting they have higher entrepreneurial ability.

5.2 Opportunity Cost Hypothesis

First-job Wage Returns to Scores We test the *opportunity cost* hypothesis by estimating the log wage equation which was specified the same as equation (1), using college graduates survey data. As shown by results from within-college regressions (columns (1)-(2) of Table 5), 1σ increase in the score is associated with 3.5% higher first-job wages.[16] This positive wage return to score supports the opportunity cost hypothesis where $b > 0$ in our conceptual framework.

We further examine how the opportunity cost increases with individual social and economic status. In Section 3, we find that the probability of firm creation is larger for those with higher status (those who are male, urban and from higher-quality high schools), suggesting that $\frac{d^2\mathbb{P}}{dSdX} < 0$ for these characteristics (X). If the opportunity cost is at work, then we should expect to observe the opportunity cost increases with the social and economic status.

We indeed find that those who are male, urban and from higher-quality high schools have larger wage-score links, further supporting the opportunity cost hypothesis. As shown in Figures 3(a)-(c) and columns (3)-(9) of Table 5, despite the limited size of our survey data, the differences in the links between score and first-job wages are large across gender and high school types. The score-wage slope for men is almost five times of that for women; that for key high school is 2.5 times of that for non-key high school; that for urban is 70% higher than that for rural.

[15] Specifically. From this, the within-college score distribution (by isolating province-year-track fixed effects and year fixed effects) is obtained , dividing the scores into five quartiles.

[16] All the results on wages have considered college sampling weight, i.e., weighted by the inverse of college sampling weight in our surveys.

State and Private Alternatives In our survey data, we also observe the sectors college graduate get into. Particularly, we are interested in the opportunity of working in the state sector. In our survey, 46% of the graduates work in the state sector (governments and state-owned-enterprises), 48% in private firms, and 5.7% becoming self-employed (including creating firms).[17] We could further examine the choices of working in the state sector or a wage job in the private sector, versus starting up businesses.

We indeed find that the state sector plays a larger part in attracting higher-score students. Using self-employment as the reference group, we report Multinomial Logit regressions results on how scores affect the relative risk of working in the state and private sectors. As shown in Appendix Table C.5, the relative risk of working in the state sector (and private firms) for students with 1σ higher score is 1.38 (and 1.25) times higher respectively. Therefore, both the state and the private sectors are relevant, but the state sector plays a relatively larger role in attracting the higher-score students[18]

Summary Overall, results in this section support the *opportunity cost* hypothesis that higher score students have a larger opportunity costs from wage jobs. In addition, we find suggestive evidence that the state sector plays a larger role than the private sector in attracting higher scores students.

5.3 Risk Attitudes Hypothesis

In this subsection, we test the *risk attitude* hypothesis. Specifically, we examine whether higher score students are more risk averse, using direct measures of risk attitudes from the 2011 wave of our college graduate survey. We have two risk attitude questions in the survey. The first question is "Do you prefer to get RMB 1,000 with certainty or get 0 and 2,000 with equal chance?" Answering yes indicates the student is risk averse. The second question is "Do you agree with the statement that ensuring certain returns to investment is more important than taking more risks to gain higher returns?" Respondents were asked to choose from a scale of 1-5, with 1 indicating strongly disagree and 5 strongly agree. In this case, a larger number means more risk averse. As expected, the two measures are positively correlated, with a correlation coefficient of 0.21.

Our test results do not support the hypothesis that higher score students are more risk averse. In regressions with the risk attitude measures as dependent variables (columns (1) and (3) of Table

[17] As described by Cook et al. (2020), college graduates in China are attracted to the state sector.

[18] The score-wage result holds if we further control for dummies indicating the state sector and the private firms, with a coefficient of 0.036, close to that in column (2) of Table 5.

6), the coefficient on the score is not statistically significant, meaning that scores cannot explain risk aversion. We also test whether the score-risk correlations vary by individual social and economic status variables in a way consistent with the variation for score-firm creation link. Results reported in columns (2) and (4) of Table 6) do not suggest such a variation. For instance, to be consistent with the stronger negative link between score and firm creation for men, the score-risk aversion relationship should also be stronger for men. We find no such evidence as the coefficient for the interaction term male*score is statistically insignificant. Interestingly, we find that women are more risk averse (in columns (2) and (4), which might serve as evidence that validate our measures of risk aversion.

Although we show that the negative link between score and firm creation cannot be explained by the risk attitude hypothesis, our findings do not belittle the importance of risk aversion and other personal traits for entrepreneurship. Even though this is not our focus, we find personal traits play important roles in firm creation. For example, we find earlier that women, rural individuals and those from lower-quality high schools are less likely to create a firm (Table 2), and these differences cannot be explained by wages. In fact, they have lower wages as shown in Table 5. This suggests that risk aversion and other personal preferences do play an important role in forming entrepreneurship.

5.4 Implications on Talent Allocation

Altogether, empirical results seem to fit better with the *opportunity cost* hypothesis, but not the *lower entrepreneurial ability* or *risk attitude* hypothesis. In fact, we find firms created by higher-score individuals are more successful, implying a positive correlation between scores and entrepreneurial ability. We also find that among the alternative opportunities, the state sector appears to be relatively more important than the private firms in absorbing these talents.

Do our findings suggest talent misallocation? To answer this question, we examine the link between score and firm creation by industry and by province. Specifically, we want to know whether the link varies by the importance of the state across industry and province. If we find a stronger negative link between score and firm creation in industries or provinces with more state penetration, it suggests that more higher-score individuals will enter the entrepreneur sector with more retreat of state intervention in the economy ; this is indeed what we find.

Estimates by Industry We plot the industry-specific estimates in Figure 6(a). Here, the dependent variable is a dummy indicating whether an individual created a firm in a certain industry (i.e., the variable is zero if he or she does not create any firm or creates a firm in other industries). To make

the estimates comparable, we plot estimates relative to the mean of the dependent variable. As shown, the estimates vary greatly: 1σ score decreases the probability of entering the construction sector more than 30% but that of science and technology service by 1%.

Our main interest is to see whether the heterogeneities across sectors also reflect the degree of state penetration in certain industries. The variation across industries, however, partly reflects the importance of human capital across industries. To take this into account for each industry, we use data on the share of state fixed investment (relative to total investment) to measure state penetration, and use the average schooling years of employees to measure human capital. After controlling for average schooling across industries, we plot the revitalized relationship between the share of state fixed investment and our estimates by industry.[19] As shown in Figure 6(b), one log point in the SOE share is associated with a -0.059 decrease in the impact or around 59% of the mean.

These results imply that higher-score individuals are more likely to avoid industries with increased state penetration.

Estimates by Province We further estimate equation (1) by province and present the estimates (relative to the provincial mean) in Figure 5(a). Again, entrepreneurial reluctance of higher-score individuals vary greatly across provinces: the negative relationship between score and firm creation is smaller in coastal provinces than that in inland provinces. For instance, the magnitude for Zhejiang (6.7%) is around one third that for Xinjiang (19.6%).

While many factors are underlying the provincial variation, we find a clear negative relationship between the province-specific estimate and the share of state investment, as shown in Figure 5(b). One log point in the SOE share is associated with a -0.04 decrease in the impact, about 40% of the mean.

One implication from this pattern is that if all provinces in China were like Zhejiang, more higher-score individuals would start businesses, which would lead to more successful firms, as shown by aforementioned evidence regarding firm success. Thus, there remains an opportunity to improve the allocation of talents to the entrepreneur sector.

[19] Because the number of firms vary greatly across industries, the estimate is weighted by the number of firms in each industry.

6 Conclusion

Despite a rich literature on entrepreneurship, evidence on the role of individual talent in firm creation is still scarce. By focusing on arguably the most important talent measure in the Chinese society, we document that within colleges, individuals with higher college entrance exam scores are *less likely* to create firms. This negative link is even stronger for the relatively more advantaged groups: men, urban students, and those from better high schools.

However, the negative relationship between scores and firm creation is unlikely because higher-score individuals have lower entrepreneurial ability. On the contrary, firms created by higher-score individuals are larger and more successful. The opposite patterns for *ex ante* firm creation and *ex post* firm success are more consistent with the hypothesis that scores are positively correlated with entrepreneurial ability but higher-score students are attracted away by non-entrepreneur sectors. Our study provides a depiction of the choice of the talented valued by the education system in China and reveals scope for improvement in talent allocation.

We are not aware of existing studies that link individual-level SAT (or other exam) score to firm creation and firm success in the U.S. or other countries.[20] We hope that our study opens new avenues for future research to compare talent allocation across countries.

References

[1] Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Victor Quintas (2021), "The Misallocation of Women's Talent Across Countries: Evidence from Personnel Data."

[2] Baumol, William J., Robert E. Litan, and Carl J. Schramm (2007), *Good Capitalism, Bad Capitalism, and the Economics of Growth and Prosperity*. Yale University Press.

[3] Bertrand, Marianne, and Antoinette Schoar (2003), "Managing with Style: The Effect of Managers on Firm Policies," *Quarterly Journal of Economics,* 118(4): 1169-1208.

[4] Blanchflower, David G., and Andrew J. Oswald (1998), "What Makes an Entrepreneur?" *Journal of Labor Economics,* 16(1): 26-60.

[5] Bloom, Nicholas, and John Van Reenen (2007), "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics,* 122(4): 1351-1408.

[6] Chevalier, Judith, and Glenn Ellison (1999), "Are Some Mutual Fund Managers Better than others? Cross?sectional Patterns in Behavior and Performance," *Journal of Finance,* 54(3): 875-899.

[20] That said, there exist studies examining SAT scores and mutual fund performance (e.g., Chevalier and Ellison 1999). However, such a measure refers to average SAT scores for colleges rather than individual-level scores and thus likely combines both college effect and individual effect.

[7] Cook, Sinclair, Hongbin Li, Lingsheng Meng, and Yanyan Xiong (2020), "Brain Drain to the State Sector: Job Preferences and Outcomes for Chinaâs College Graduates", working paper.

[8] Evans, David S., and Boyan Jovanovic (1989), "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy,* 97 (4): 808-827.

[9] Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2013), "Human Capital and Regional Development," *Quarterly Journal of Economics,* 128: 105-164.

[10] Hayward, Mathew LA, Dean A. Shepherd, and Dale Griffin (2006), "A Hubris Theory of Entrepreneurship," *Management Science,* 52(2): 160-172.

[11] Heckman, James (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement,* 5(4): 475-492.

[12] Hsieh, ChangâTai, Erik Hurst, Charles I. Jones, and Peter J. Klenow (2019), "The Allocation of Talent and U.S. Economic Growth," *Econometrica,* 87(5): 1439-1474.

[13] Hsieh, Chang-Tai, and Peter J. Klenow (2009), "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics,* 124(4): 1403-1448.

[14] Iyigun, Murat F., and Ann L. Owen (1998), "Risk, Entrepreneurship, and Human-Capital Accumulation," *American Economic Review,* 88(2): 454-457.

[15] Jia, Ruixue and Hongbin Li (2019), "Just Above the Cutoff: The Return to Elite Education in China," Working Paper.

[16] Lazear, Edward (2004), "Balanced Skills and Entrepreneurship," *American Economic Review,* 94(2): 208-211.

[17] Levine, Ross, and Yona Rubinstein (2017), "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *Quarterly Journal of Economics,* 132(2): 963-1018.

[18] Levine, Ross, and Yona Rubinstein (2019), "Selection into Entrepreneurship and Self-Employment" NBER Working Paper No. w25350.

[19] Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1991), "The Allocation of Talent: Implications for Growth," *Quarterly Journal of Economics,* 106(2): 503-530.

[20] Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz (2012), "The Short-and Long-term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics,* 4(1): 1-29.

[21] Parker, Simon (2018), *The Economics of Entrepreneurship*. Cambridge University Press.

[22] Syverson Chad (2011), "What Determines Productivity?" *Journal of Economic Literature,* 49(2): 326-365.

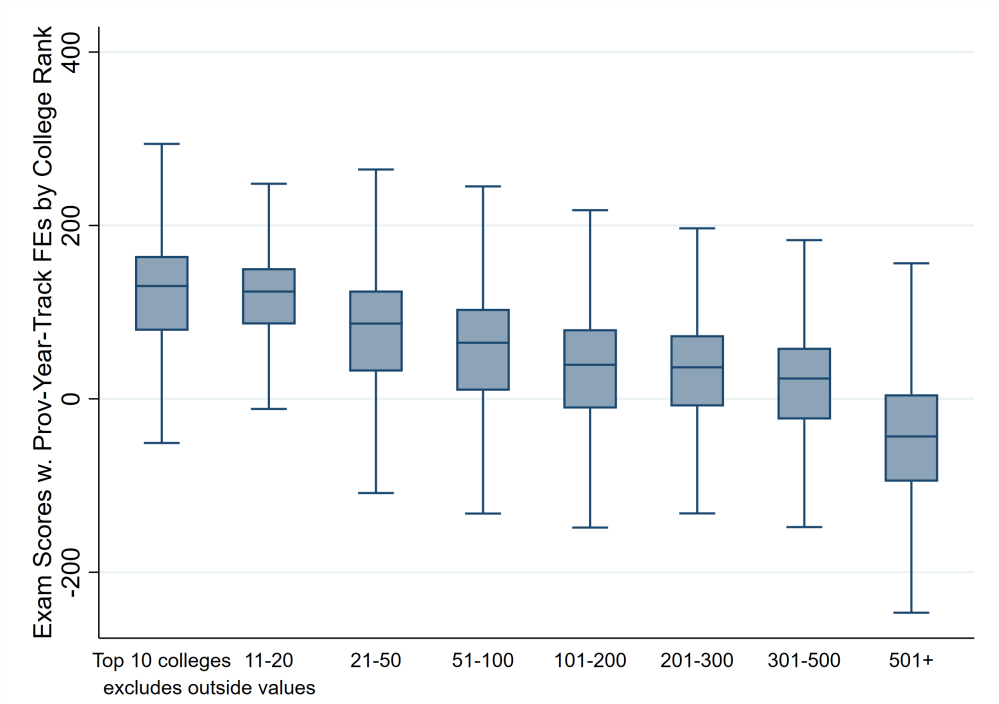
[23] Ye, Hua (2015), "Key-point Schools and Entry into Tertiary Education in China, *Chinese Sociological Review* 47(2): 128-153.

[24] Young, Alwyn (2000), "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China, *Quarterly Journal of Economics,* 115(4): 1091-1135.

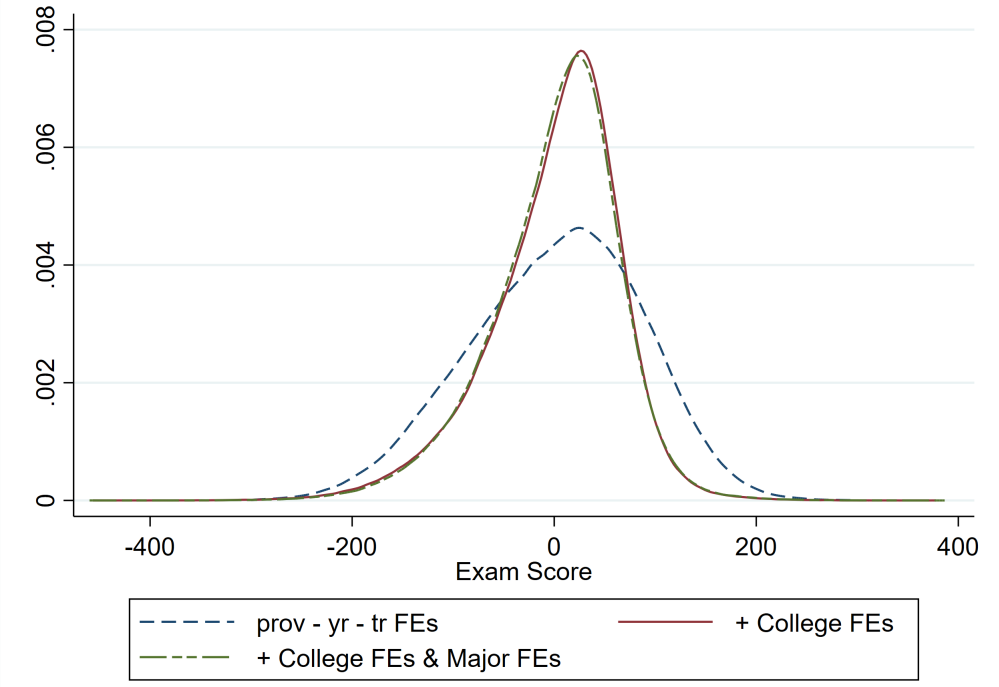
[25] Zhao, Yong (2009), *Catching Up or Leading the way: American Education in the Age of Globalization*. ASCD.

Figure 1: Exam Scores Distribution across and within Colleges

(a) Exam Scores by College Ranks



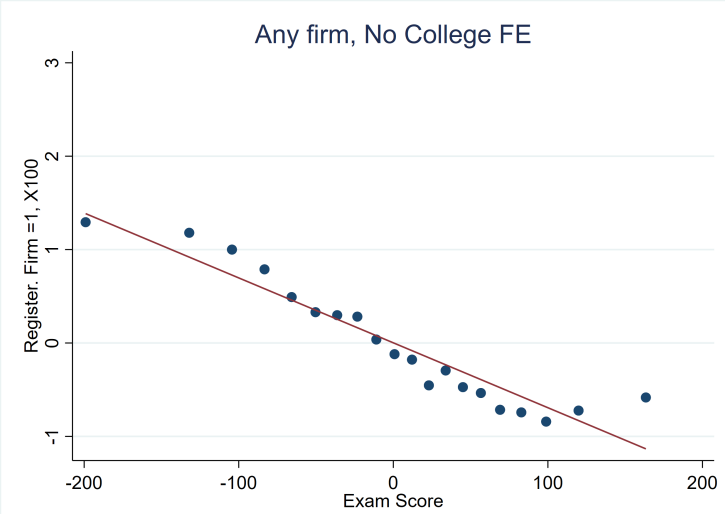
(b) Exam Score Distributions



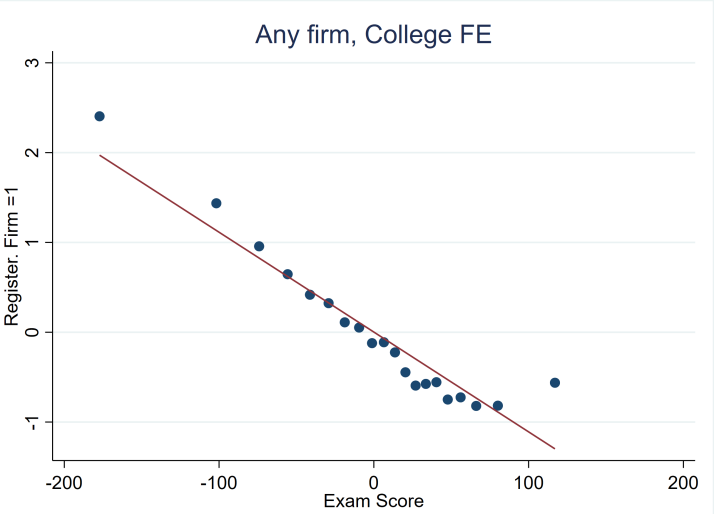
Notes: Figure (a) shows that better colleges have higher-score students on average. Meanwhile, score distributions overlap across different batches of colleges. Figure (b) confirms the wide variation within colleges.

Figure 2: College Entrance Exam Scores vs. Firm Creation

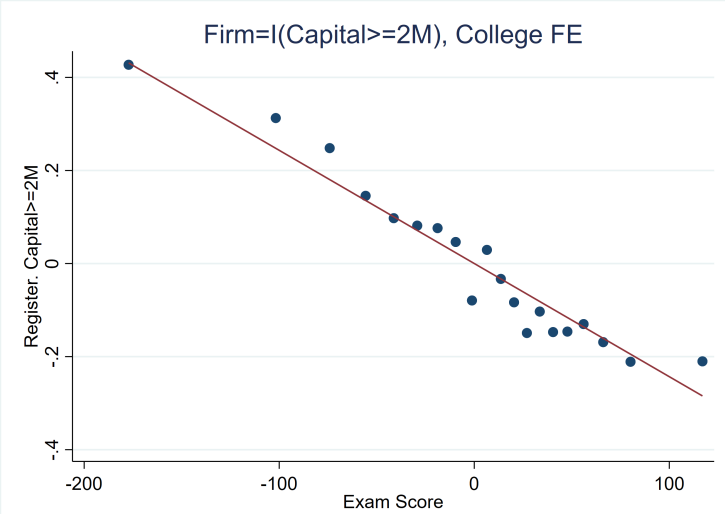
(a) No college FE, any firm



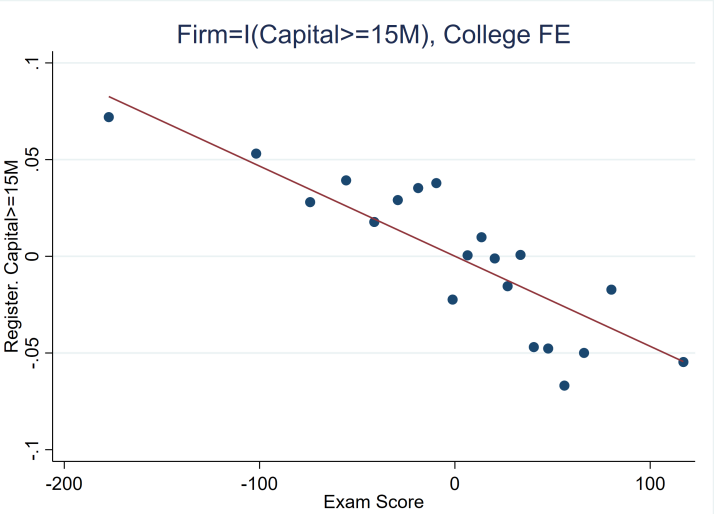
(b) College FE, any firm



(c) College FE, registered capital ≥2M



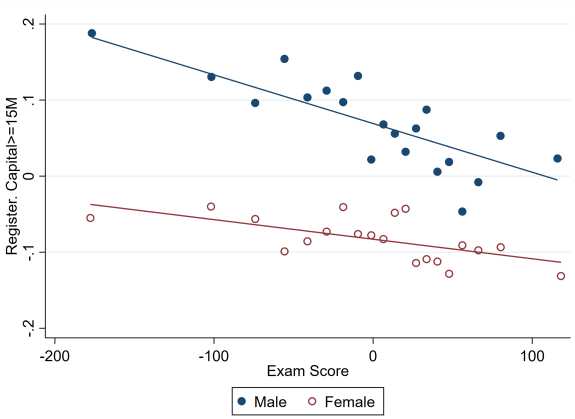
(d) College FE, registered capital ≥15M



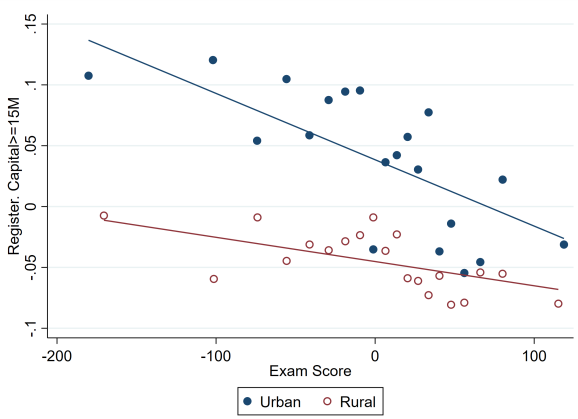
Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million.

Figure 3: Academic Talent and Firm Creation (Within Colleges), by Personal Background

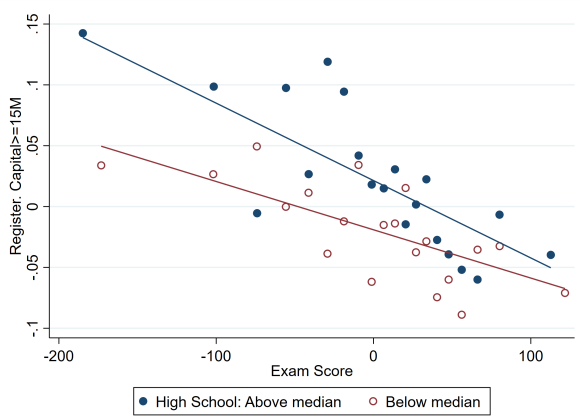
(a) Male vs. Female



(b) Urban vs. Rural



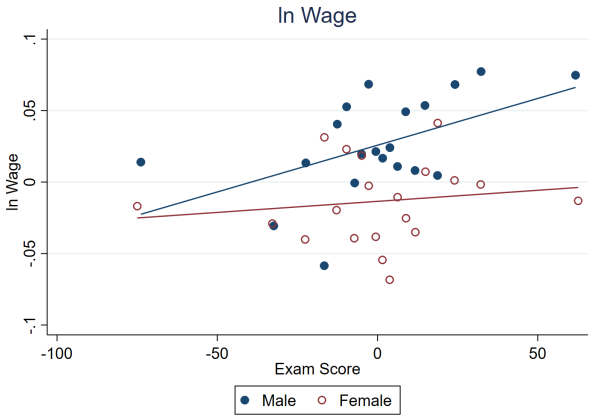
(c) Better vs. Worse High Schools



Notes: This figure shows that the negative relationship between scores and firm creation is even stronger for male, urban individuals, and those from better high schools. Province-year-track FEs and college FEs are controlled in all figures.
We focus on large firms (those with registered capital no less than RMB 15 million) in these figures and show similar patterns when varying firm definition in Table 3.

Figure 4: Academic Talent and First-Job Wages (Within Colleges), by Personal Background

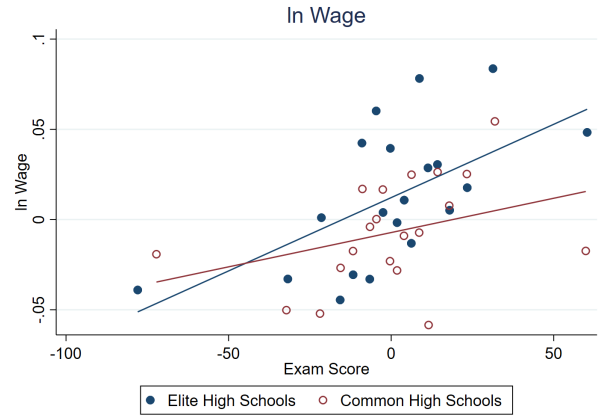
(a) Male vs. Female



(b) Urban vs. Rural



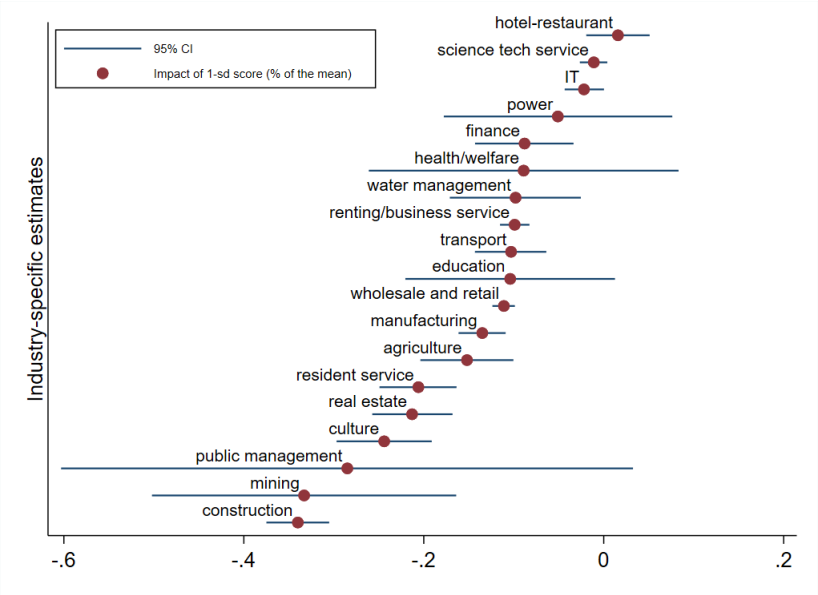
(c) Better vs. Worse High Schools



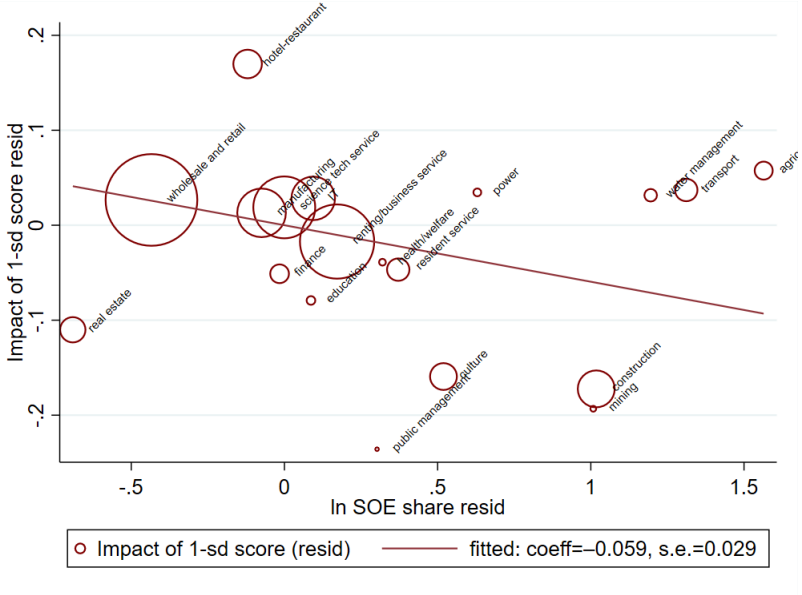
Notes: This figure shows that the positive score-wage link is stronger for male, urban individuals, and those from better high schools. Province-year-track FEs and college FEs are controlled in all figures. This data comes from the College Student Survey we conducted during 2010-2015 (see Jia and Li 2019 for more discussion). The wage refers to the first job wages.

Figure 5: Score-Firm Relationships by Industry

(a) Estimates By Industry



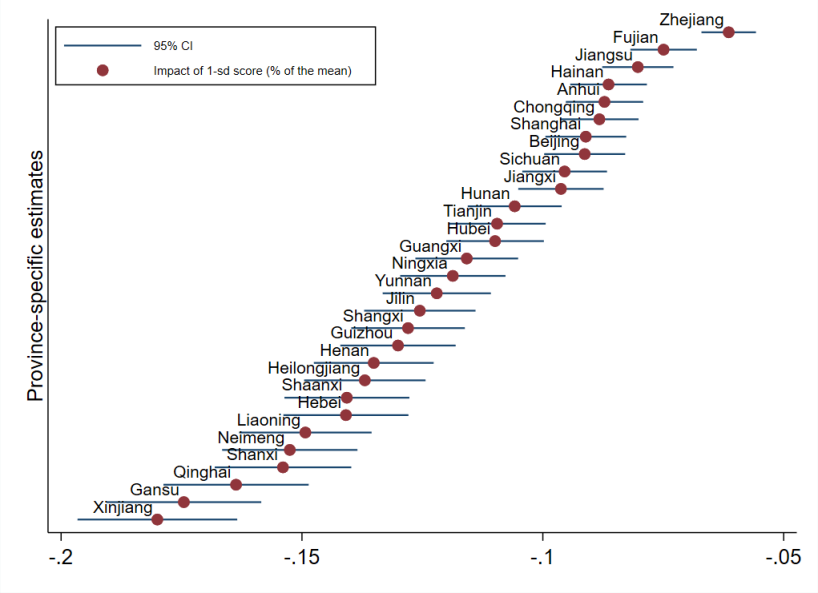
(b) Estimates vs. SOE share



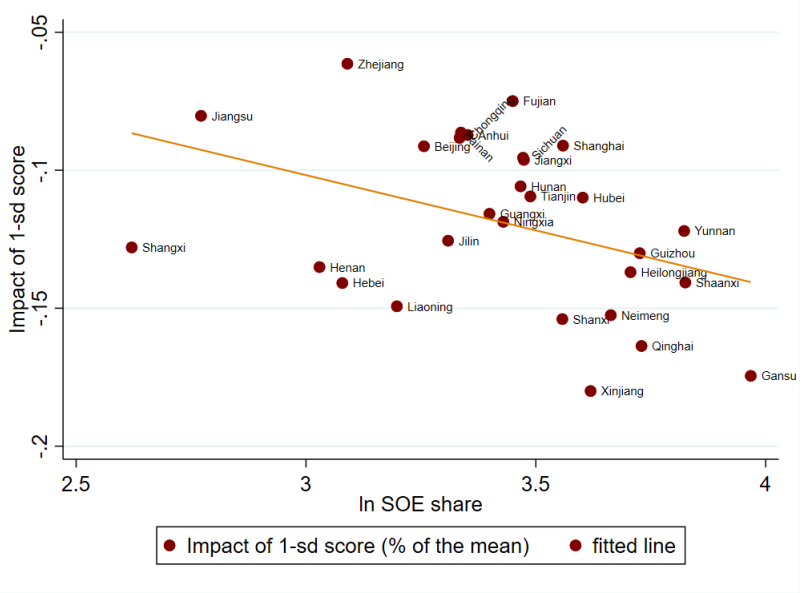
Notes: Figure (a) plots the estimates by industry where the dependent variable is a dummy indicating entering a certain industry. The coefficients can be interpreted as "one standard deviation score is negatively associated with X% of the dependent variable". Figure (b) plots the correlation between these estimates and the share of SOE investment by industry, where we control for the average years of schooling of employees by industry. The size of the circles indicates the number of firms in each industry.

Figure 6: Score-Firm Relationships by Province

(a) Estimates by Province



(b) Estimates vs. SOE share



Notes: Figure (a) plots the estimates by province. The coefficients can be interpreted as "one standard deviation score is negatively associated with X% of the dependent variable". Figure (b) plots the correlation between these estimates and the share of SOE investment by province.

Table 1: Summary Statistics

A. Information on Individuals	Mean	Std Dev	p25	p75	Obs.
Prob of Creating a Firm (*100)					
Any Firm	7.21	25.87			1,814,501
Firms w. Capital>=RMB 2M	1.65	12.73			1,814,501
Firms w. Capital>=RMB 15M	0.40	6.29			1,814,501
					1,814,501
College Entrance Exam Score	436.9	101.4	371	507	1,814,501
Exam Score (w. prov-yr-tr FEs)	0.0	88.9	-58	63	1,814,501
Exam Score (w. prov-yr-tr FEs+college FEs)	0.0	68.4	-35	44	1,814,501
Exam Score (w. prov-yr-tr FEs+college FEs + major FEs)	0.0	66.7	-35	43	1,814,501
Birth Year	1982.7	1.7	1982	1984	1,814,501
Male	0.54	0.50			1,814,501
Urban	0.53	0.50			1,814,501
High School (Sh. Stud. in Top 100 Colleges)	0.14	0.13	0.048	0.194	1,814,501
B. Information on Firms	Mean	Std Dev	p25	p75	Obs.
Registered Capital	632	6042	20	200	170,087
ln Registered Capital	4.28	1.74	3.04	5.30	170,087
Out of Home Province	0.39	0.49			170,087
Expanding (Investing in Other Firms)	0.05	0.22			170,087
Becoming Listed (*1000)	0.36	18.94			170,087
Establishment Year	2010.6	3.69	2009	2013	169,668
C. Firms by Industry	Percent				
Wholesale and Retail	30.91				
Leasing and business services	20.19				
Scientific research and tech services	13.96				
Manufacturing	8.54				
Info. Trans, software and i.t. services	6.84				
the other 15 industries	19.56				

Table 2: Exam Scores and Firm Creation, within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Any Firm			Firm=I(Capital≥2M)			Firm=I(Capital≥15M)	
mean		7.12			1.64			0.40	
Exam score (sd)	-0.756	-0.770	-0.728	-0.165	-0.174	-0.173	-0.032	-0.035	-0.039
	(0.032)	(0.032)	(0.032)	(0.011)	(0.012)	(0.012)	(0.005)	(0.005)	(0.005)
Male		3.402	3.538		0.850	0.866		0.184	0.184
		(0.063)	(0.064)		(0.024)	(0.025)		(0.011)	(0.011)
Urban		0.278	0.218		0.239	0.226		0.094	0.087
		(0.049)	(0.049)		(0.022)	(0.022)		(0.011)	(0.011)
High school quality (sd)		0.260	0.273		0.131	0.129		0.045	0.044
		(0.032)	(0.031)		(0.014)	(0.014)		(0.007)	(0.007)
ln GDP per capita		0.095	0.091		0.039	0.038		0.012	0.011
(birth county, 2001)		(0.031)	(0.030)		(0.014)	(0.014)		(0.006)	(0.006)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE		Y	Y		Y	Y		Y	Y
major FE			Y			Y			Y
Obs. (excl singletons)	1,814,488	1,814,482	1,799,638	1,814,488	1,814,482	1,799,638	1,814,488	1,814,482	1,799,638

Notes: Conditional on colleges, one's exam score is strongly and negatively associated with firm creation. This finding is robust to defining firms by size and considering personal background and major fixed effects.
Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. Standard errors in the paraphrases are clustered at the college level.

Table 3: Exam Scores and Firm Creation by Personal Background, within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Any firm (mean 7.12)			Firm=I(Capital>=2M) (mean:1.65)			Firm=I(Capital>=15M) (mean:0.40)		
Exam score (sd)	-0.549	-0.649	-0.640	-0.140	-0.123	-0.150	-0.026	-0.019	-0.030
	(0.030)	(0.040)	(0.033)	(0.012)	(0.015)	(0.013)	(0.006)	(0.007)	(0.006)
Male×Score	-0.358			-0.049			-0.014		
	(0.040)			(0.019)			(0.008)		
Urban×Score		-0.134			-0.054			-0.016	
		(0.042)			(0.019)			(0.009)	
Above-med Highsch×Score			-0.276			-0.057			-0.014
			(0.044)			(0.021)			(0.009)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Male×college FE	Y			Y			Y		
Urban×college FE		Y			Y			Y	
Above-med Highsch×col FE			Y			Y			Y
Obs. (excl singletons)	1,814,454	1,814,448	1,814,456	1,814,454	1,814,448	1,814,456	1,814,454	1,814,448	1,814,456

Notes: This table shows that the negative score-firm creation link is stronger for men, urban individuals, and those from better high schools. These patterns are robust to defining firms by capital size. Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million. Standard errors in the paraphrases are clustered at the college level.

Table 4: Exam Scores and Firm Success, within Colleges
Administrative Data, Firm-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Validity of the instrument		**Selection**	**A. Heckman two-step estimates**			
Dependent Var.	$peer_{c,t}$ (sd)		marginal effect	ln Reg Capital within ind	Out of Home Prov	Expand	Listed (*1000)
mean dependent var				4.281	0.394	0.049	0.359
Exam score (sd)	0.0267	0.0006		0.0061	0.0363	0.0051	0.0686
	(0.0048)	(0.0004)		(0.0039)	(0.0010)	(0.0005)	(0.0448)
$peer_{c,t}$ (sd)			0.0238				
			(0.0044)				
λ (inverse mills ratio)				0.0944	-0.0044	-0.0747	-3.6334
				(0.2627)	(0.0709)	(0.0349)	(3.0466)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y
Other personal controls	Y	Y	Y	Y	Y	Y	Y
college FE		Y	Y	Y	Y	Y	Y
Obs. (excl singletons)	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074
				B. Estimates conditional on entry			
				ln Reg Capital within ind	Out of Home Prov	Expand	Listed (*1000)
Exam score (sd)				0.0068	0.0375	0.0053	0.0712
				(0.0044)	(0.0017)	(0.0005)	(0.0398)
prov-track-year FE				Y	Y	Y	Y
age FE				Y	Y	Y	Y
Other personal controls				Y	Y	Y	Y
college FE				Y	Y	Y	Y
Obs. (excl singletons)				169,332	169,349	169,349	169,349

Notes: The table shows that one's score is positively associated with firm success. To address the concern of endogenous entry, we employ the entrepreneurial propensity across provincial origins of one's peers within the same college as an instrument. Columns (1)-(2) show that while one's own score is correlated with the instrument, it is not the case any more within a college-cohort. Column (3) shows that the instrument is indeed a strong predictor for firm creation. Panel A reports the Heckman two-step estimates and Panel B the estimates conditional on entry.
Other personal controls include gender, rural status, higher school types and log GDP per capita of one's birth county. Standard errors in the paraphrases in Panel B are clustered at the college level.

Table 5: Exam Scores and First-job Wages by Personal Background, within Colleges
Survey Data, Individual-level Analysis

Dependent var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ln Wage		ln Wage					
Exam score (sd)	0.030	0.037	0.009	0.011	0.019	0.020	0.021	0.022
	(0.010)	(0.012)	(0.016)	(0.016)	(0.014)	(0.015)	(0.015)	(0.015)
Male×Score			0.039	0.037				
			(0.025)	(0.025)				
Urban×Score					0.016	0.015		
					(0.018)	(0.017)		
Elite Highsch×Score							0.034	0.035
							(0.020)	(0.020)
Male		0.040	0.042					
		(0.017)	(0.012)					
Urban		0.029			0.024			
		(0.010)			(0.008)			
Elite High School		0.012					0.017	
		(0.010)					(0.011)	
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y
Male×college FE				Y				
Urban×college FE						Y		
Elite Highsch×col FE								Y
Obs. (excl singletons)	14,697	11,164	14,697	14,697	14,142	14,140	11,535	11,534

Notes: This table shows that scores are positively associated with first-job wages, and the positive score-wage link is stronger for male, urban individuals, and those from better high schools. This data comes from the College Student Survey we conducted during 2010-2015. Standard errors in the paraphrases are clustered at the college level.

Table 6: Exam Scores and Risk Aversion, within Colleges
Survey Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	Not taking lottery, 0/1			Important to ensure certain return in investment,1-5		
Mean	0.64			3.09		
Exam score (sd)	0.003	0.015	0.079	-0.005	0.036	-0.105
	(0.039)	(0.032)	(0.071)	(0.057)	(0.057)	(0.106)
Male×Score			-0.002			0.165
			(0.068)			(0.134)
Urban×Scor			-0.056			0.181
			(0.072)			(0.163)
Elite High Sch×Scor			-0.049			-0.072
			(0.069)			(0.102)
Male		-0.055	-0.054		-0.105	-0.110
		(0.025)	(0.025)		(0.045)	(0.046)
Urban		0.027	0.028		0.010	0.005
		(0.026)	(0.026)		(0.051)	(0.049)
Elite High school		-0.012	-0.011		-0.086	-0.085
		(0.018)	(0.018)		(0.033)	(0.032)
prov-track-year FE	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y
Observations	3,134	3,001	3,001	3,090	2,961	2,961

Notes: Notes: This table shows that there is no strong relationship between scores and risk aversion.
This data comes from the College Student Survey we conducted in 2011. We did not cover risk questions in the other years. Standard errors in the paraphrases are clustered at the college level.

Online Appendix

Table of Contents

A Correlations b/w Scores and Personal Characteristics **A-2**

B Score-Firm Creation: Additional Results **A-2**

B.1 Including Major Fixed Effects A-3

B.2 Results Excluding Province-Years with Missing IDs A-4

B.3 Results Excluding Repeated Exam Takers A-5

B.4 Results Excluding Students with Scores on the Tails A-6

B.5 Results by Majors A-7

B.6 Results By College Ranks A-8

C Interpretations: Additional Results **A-9**

C.1 More Results on IV Validity A-9

C.2 Visualization: Score-Firm Success Conditional on Entry A-9

C.3 Checking Exclusion Restriction A-11

C.4 A Survival Analysis A-12

C.5 The State Sector and the Private Firms as Alternatives A-12

A Correlations b/w Scores and Personal Characteristics

As discussed, exam scores capture both ability and family background. We examine the correlations between one's score and his or her personal characteristics in Table A. Within college, females and males do not differ in the *gaokao* score, though men do better across colleges. Rural students have higher scores, both across and within colleges, probably because only the better rural students could attend high schools and participate in the exam. Perhaps not surprisingly, students from better high schools and richer counties tend to have higher scores, both across and within colleges.

Table A: Exam Scores and Personal Characteristics, across and within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)
	Exam Score	
Mean	443.8	
Male	4.156	-0.104
	(0.651)	(0.337)
Urban	-22.439	-20.387
	(0.669)	(0.491)
High school quality (sd)	38.361	16.112
	(0.811)	(0.534)
ln GDP per capita	1.078	0.928
(birth county, 2001)	(0.266)	(0.170)
age FE	Y	Y
prov-track-year FE	Y	Y
college FE		Y
R-squared	0.365	0.570
Obs. (excl singletons)	1,814,495	1,814,482

Notes: Standard errors in the paraphrases are clustered at the college level.

B Score-Firm Creation: Additional Results

B.1 Including Major Fixed Effects

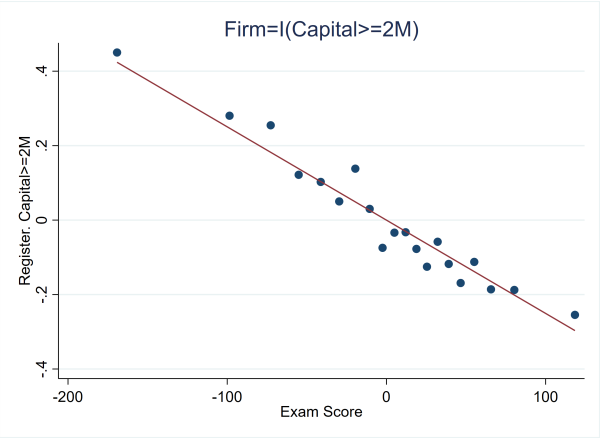
Figure B.1 shows that the negative link between exam scores and firm creation within colleges hold after controlling for 12-major fixed effects. Because major choices reflect one's own preferences, we focus on with-college comparison and report results without and with major fixed effects.

Figure B.1: Exam Scores and Firm Creation: With College FEs and Major FEs

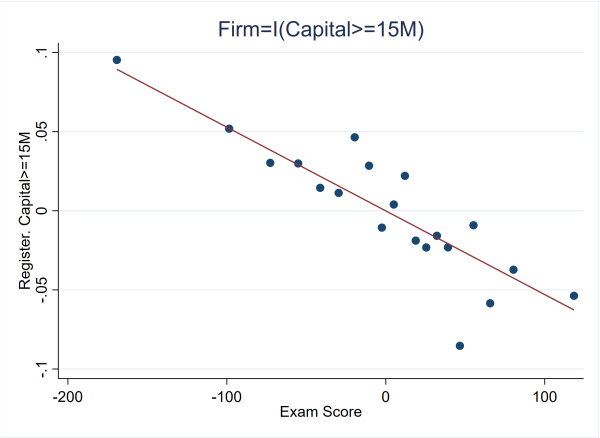
(a) With College FE and Major FE, any firm



(b) With College FE and Major FE, reg. capital≥2M



(c) With College FE and Major FE, reg. cap≥15M



We report the relationship between exam scores and firm creation by majors in Table B.5.

B.2 Results Excluding Province-Years with Missing IDs

Our findings remain very similar if we restrict data to individuals in province-years whose id missing rates are lower than 5% (Panel A of Table B.2) or even 1% (Panel B of Table B.2.)

Table B.2: Province-Years with Little Missing Info., within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Firm		Firm=I(Capital≥2M)		Firm=I(Capital≥15M)	
prov-track-year FE	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y
age FE and other personal controls		Y		Y		Y
	A. sample: missing id < 5%					
Exam score (sd)	-0.762	-0.766	-0.163	-0.171	-0.029	-0.032
	(0.043)	(0.043)	(0.016)	(0.016)	(0.007)	(0.007)
Obs. (excl singletons)	943,383	943,377	943,383	943,377	943,383	943,377
	B. sample missing id < 1%					
Exam score (sd)	-0.932	-0.908	-0.221	-0.215	-0.037	-0.037
	(0.100)	(0.109)	(0.039)	(0.042)	(0.018)	(0.018)
Obs. (excl singletons)	428,412	428,403	428,412	428,403	428,412	428,403

Notes: Other personal controls include gender, rural status, higher school types and log GDP per capita of one's birth county. Standard errors in the paraphrases are clustered at the college level.

B.3 Results Excluding Repeated Exam Takers

Our findings remain similar after excluding repeated exam takers, as shown in Table B.3.

Table B.3: Excluding Repeated Exam Takers, within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Firm		Firm=I(Capital≥2M)		Firm=I(Capital≥15M)	
Exam score (sd)	-0.729	-0.752	-0.166	-0.175	-0.034	-0.038
	(0.034)	(0.035)	(0.013)	(0.013)	(0.006)	(0.006)
Female		-3.451		-0.861		-0.185
		(0.067)		(0.028)		(0.013)
Rural		-0.275		-0.257		-0.101
		(0.058)		(0.025)		(0.013)
High school quality (sd)		0.284		0.139		0.050
		(0.034)		(0.016)		(0.008)
ln GDP per capita		0.101		0.047		0.017
(birth county, 2001)		(0.034)		(0.016)		(0.007)
prov-track-year FE	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y
age FE		Y		Y		Y
Obs. (excl singletons)	1,310,541	1,310,508	1,310,541	1,310,508	1,310,541	1,310,508

Notes: Standard errors in the paraphrases are clustered at the college level.

B.4 Results Excluding Students with Scores on the Tails

Our baseline finding holds if we restrict data to individuals with scores between 10th-90th percentiles on the score distribution within a college. As shown in Table B.4, the estimates become even stronger if we exclude the students on the tails.

Table B.4: Excluding Students with Scores on the Tails, within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	Any Firm		Firm=I(Capital≥2M)		Firm=I(Capital≥15M)	
Exam score (sd)	-0.869	-0.924	-0.199	-0.222	-0.041	-0.050
	(0.050)	(0.052)	(0.019)	(0.019)	(0.008)	(0.008)
Female		-3.314		-0.843		-0.182
		(0.066)		(0.027)		(0.012)
Rural		-0.248		-0.226		-0.085
		(0.054)		(0.024)		(0.012)
High school quality (sd)		0.334		0.146		0.050
		(0.033)		(0.015)		(0.008)
ln GDP per capita		0.089		0.048		0.012
(birth county, 2001)		(0.035)		(0.015)		(0.007)
prov-track-year FE	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y
age FE		Y		Y		Y
Obs. (excl singletons)	1,466,648	1,466,622	1,466,648	1,466,622	1,466,648	1,466,622

Notes: Standard errors in the paraphrases are clustered at the college level.

B.5 Results by Majors

Our baseline finding holds when examining the pattern by major, as shown by Table B.5.

Table B.5: Exam Scores and Firm Creation by Majors, within Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent var.		Any firm			Firm=I(capital≥2M)			Firm=I(capital≥15M)	
	STEM	Econ-Fin-Law	Humanity	STEM	Econ-Fin-Law	Humanity	STEM	Econ-Fin-Law	Humanity
Mean Dependent var.	6.50	8.21	7.52	1.52	2.02	1.41	0.35	0.55	0.28
Exam score (sd)	-0.574	-0.625	-0.890	-0.163	-0.208	-0.157	-0.044	-0.052	-0.021
	(0.041)	(0.055)	(0.054)	(0.017)	(0.025)	(0.019)	(0.007)	(0.012)	(0.009)
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
other personal controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs. (excl singletons)	928,984	498,634	371,931	928,984	498,634	371,931	928,984	498,634	371,931

Notes: Other personal controls include gender, rural status, higher school types and log GDP per capita of one's birth county. Standard errors in the paraphrases are clustered at the college level.

B.6 Results By College Ranks

Our baseline finding holds when examining the pattern by college rank, as reported in Table B.6.

Table B.6: Exam Scores and Firm Creation by College Rank, with Colleges
Administrative Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Any Firm			Firm=I(Capital≥2M)			Firm=I(Capital≥15M)	
mean	8.33	7.69	7.1	2.27	1.98	1.57	0.6	0.49	0.38
college rank	Top 10	11-100	100+	Top 10	11-100	100+	Top 10	11-100	100+
Exam score (sd)	-0.773	-1.232	-0.624	-0.188	-0.304	-0.149	-0.045	-0.067	-0.029
	(0.222)	(0.085)	(0.031)	(0.086)	(0.031)	(0.012)	(0.034)	(0.016)	(0.005)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
other personal controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs. (excl singletons)	34,761	266,670	1,500,351	34,761	266,670	1,500,351	34,761	266,670	1,500,351

Notes: Other personal controls include gender, rural status, higher school types and log GDP per capita of one's birth county. Standard errors in the paraphrases are clustered at the college level.

C Interpretations: Additional Results

C.1 More Results on IV Validity

If our instrument is valid, we should see no strong correlations between our instrument and scores and other personal characteristics. This is true within colleges. As reported in column (1) of Table C.1, without controlling for college sorting, personal characteristics are correlated with our peer exposure measure. But such correlations are close to zero within colleges (column (2)).

Table C.1: Validity Checks of the Instrument

	(1)	(2)
	$\text{peer}_{c,t}$ (sd)	
Exam score (sd)	0.0267	0.0006
	(0.0048)	(0.0004)
Male	0.0042	-0.0004
	(0.0038)	(0.0003)
Urban	0.0595	-0.0002
	(0.0045)	(0.0003)
High school quality (sd)	0.0116	-0.0013
	(0.0052)	(0.0003)
ln GDP per capita	0.0066	-0.0005
(birth county, 2001)	(0.0014)	(0.0002)
prov-track-year FE	Y	Y
age FE	Y	Y
college FE		Y
Observations	1,852,993	1,852,993
R-squared	0.736	0.987

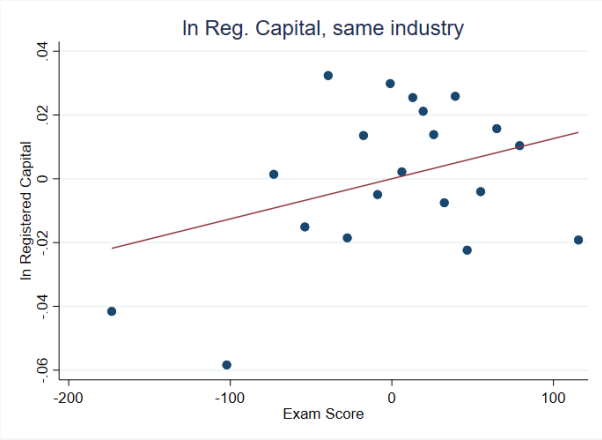
Notes: Standard errors in the paraphrases are clustered at the college level.

C.2 Visualization: Score-Firm Success Conditional on Entry

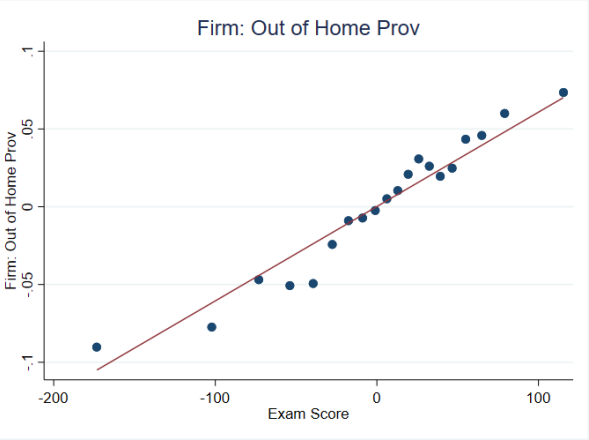
Given entry, exam scores (within colleges) appear positively associated with multiple measures of firm success, including (a) registered capital (demeaned by industry mean), (b) the propensity of creating a firm outside one's home province, (c) the propensity of investing in other firms as a shareholder, and (d) the propensity of becoming publicly listed. Figure C.2 visualizes these correlations.

Figure C.2: Firm Capital Size and Other Success Measures (given entry)

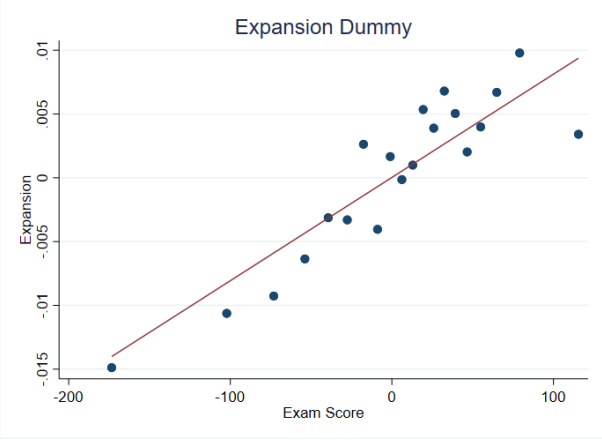
(a) log Registered Capital



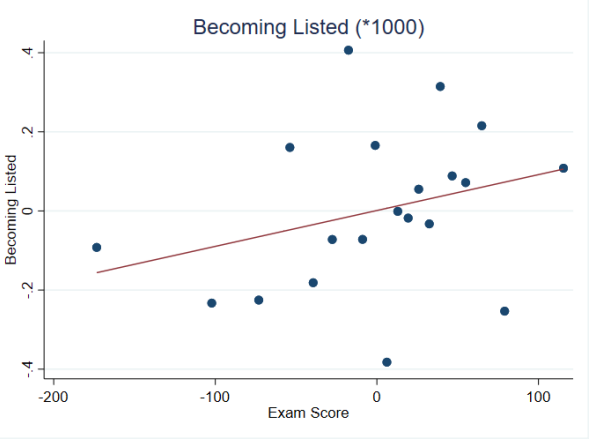
(b) Firms Created Out of Home Provinces



(c) Expansion



(d) Becoming Listed



However, we need to deal with the challenge of selection into entry when evaluating how scores affect firm. As discussed in the paper, we exploit peer composition across cohorts within colleges as an instrument to predict firm entry and use the Heckman two-step method to correct possible selection bias.

C.3 Checking Exclusion Restriction

We exploit peer composition to predict firm entry. To check whether it is a concern that peer composition also matters for firm success, we separate the peers into two groups. Specifically, we divide students' home provinces into Northern China and Southern China. Then, we use peer compositions from both the same part and the other part of China to predict entry. If peer composition matters for success, we would to see a larger impact on success when using the former as an instrument because it is more likely to help with firm success. This is not the case, as presented in Table C.3.

Table C.3: Exam Scores and Firm Success, within Colleges
Administrative Data, Firm-level Analysis

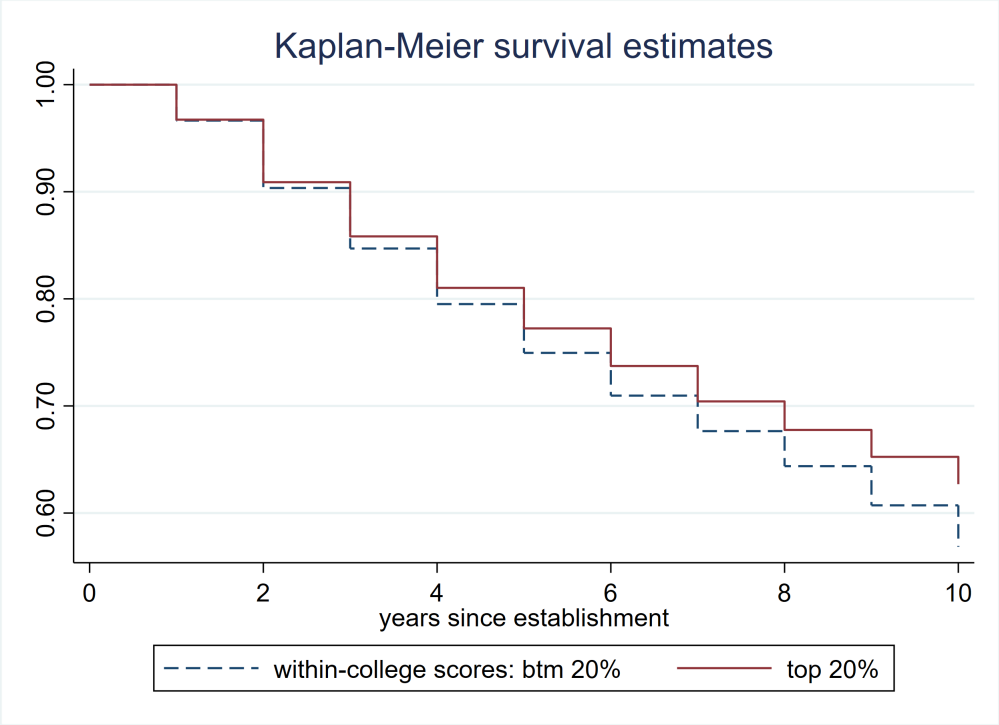
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Var.	ln Reg capital (within ind)		Out of home province		Expand		Listed (*1000)	
instrument: peer composition	Same region	Different region	Same region	Different region	Same region	Different region	Same region	Different region
Gaokao score (sd)	0.0057	0.0057	0.0353	0.0350	0.0051	0.0050	0.0742	0.0739
	(0.0039)	(0.0039)	(0.0060)	(0.0024)	(0.0005)	(0.0005)	(0.0448)	(0.0448)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y	Y
Other personal controls	Y	Y	Y	Y	Y	Y	Y	Y
Obs. (excl singletons)	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074	1,858,074

Notes: Other personal controls include gender, rural status, higher school types and log GDP per capita of one's birth county. Standard errors in the paraphrases are clustered at the college level.

C.4 A Survival Analysis

Figure C.4 shows that the firms founded by those with higher scores are more likely to survive, and the difference becomes more important over time. The estimated hazard ratio with respect to within-college exam scores is around 0.975, with a standard error of 0.006, which is consistent with the pattern in the figure.

Figure C.4: Within-college Scores vs. Firm Survival



C.5 The State Sector and the Private Firms as Alternatives

We examine how scores affect the relative risk of working in different sectors in C.5, using self-employment as the reference group. As shown in columns (1)-(2), scores are positively associated with a higher risk of working in the state sector as well as the private firms, while the former risk ratio is higher. In addition, the return to scores also appears larger in the state sector (columns (3)-(4)).

Table C.5: The First-Job Sectors, within Colleges
Survey Data, Individual-level Analysis, Reference: Self-Employment

	(1)	(2)	(3)	(4)
	Relative risk		ln Wage	
	State Sector	Private Firms	State Sector	Private Firms
Exam score (sd)	1.523	1.383	0.046	0.023
	(0.201)	(0.205)	(0.016)	(0.017)
prov-track-year FE	Y	Y	Y	Y
college FE	Y	Y	Y	Y
age FE	Y	Y	Y	Y
other personal controls	Y	Y	Y	Y
Obs. (excl singletons)	11,071	11,071	4,845	5,492

Notes: Other personal controls include gender, rural status, and higher school types. Standard errors in the paraphrases are clustered at the college level.