NBER WORKING PAPER SERIES

ENTREPRENEURIAL RELUCTANCE:
TALENT AND FIRM CREATION IN CHINA

Chong-En Bai
Ruixue Jia
Hongbin Li
Xin Wang

Working Paper 28865
http://www.nber.org/papers/w28865

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2021

We thank Ying Bai, Ting Chen, Julie Cullen, Gordon Dahl, Alexia Delfino, Hanming Fang, Roger Gordon, Josh Graff Zivin, David Mckenzie, Yona Rubinstein, Imran Rasul, Andrei Shleifer, John Van Reenen, Michael Song, Yang Song, Noam Yuchtman, Xiaobo Zhang, and participants of seminars at CEFPA Seminar, CEPR/LEAP, Fudan, Georgetown-World Bank, HKUST, LSE, Luohan Academy, NBER, Nottingham, Peking University, SITE Stockholm, Stanford China Center, UBC, UCL, UCSD, and USC for their comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Entrepreneurial Reluctance: Talent and Firm Creation in China
Chong-En Bai, Ruixue Jia, Hongbin Li, and Xin Wang
NBER Working Paper No. 28865
May 2021
JEL No. H11,J24,O12,O15

ABSTRACT

The theoretical literature has long noted that talent can be used in both the entrepreneurial and non-entrepreneurial sectors, and its allocation depends on the reward structure. We test these hypotheses by linking administrative college admissions data for 1.8 million individuals with the universe of firm registration records in China. Within a 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, but, when they do, their firms are more successful than those of their lower-score counterparts. Additional survey data suggest that higher-score individuals enjoy higher wages and are more likely to join the state sector. Moreover, the score-to-firm creation relationship varies greatly across industry, according to the size of the state sector. These findings suggest that the score is positively associated with both entrepreneurial ability and wage-job ability but higher-score individuals are attracted away by wage jobs, particularly those of the state sector.

Chong-En Bai
School of Economics and Management
Tsinghua University
Beijing 100084
China
baichn@sem.tsinghua.edu.cn

Ruixue Jia
School of Global Policy and Strategy
University of California at San Diego
9500 Gilman Drive #0519
La Jolla, CA 92093
and CEPR
and also NBER
rxjia@ucsd.edu

Hongbin Li
Stanford Institute for Economic Policy Research
366 Galvez St
Stanford
Stanford, CA 94305
lihongbinsem@gmail.com

Xin Wang
CUHK
wangx2.04@gmail.com

1 Introduction

Allocating talent to the entrepreneurial sector, vis-à-vis the rent-seeking sector, is key to sustaining economic growth (Baumol 1990; Murphy, Shleifer and Vishny 1993; Acemoglu 1995; Baumol, Litan and Schramm 2007). As noted in the seminal work of Murphy, Shleifer and Vishny (1991), however, talented people may choose to become rent-seekers rather than entrepreneurs, depending on the reward structure of a society. They point out that talent is general and can be used in both the entrepreneurial and rent-seeking sectors and that it will be allocated to rent-seeking in societies in which government officials or winners in a profession can collect large amounts of rent. While this important theory of talent allocation has been developed for three decades, few empirical studies have directly tested it. One reason is that it is challenging to obtain a dataset that includes good information on both talent and entrepreneurship. Additionally, the data needs to include a large sample size, given the small probability of firm creation. It is also difficult to separate the impact of talent from that of education and family background.

In this paper, we study whether talented people are more or less likely to become entrepreneurs and test whether this relationship is consistent with the talent allocation theory, in which entrepreneurial and rent-seeking sectors compete for the same talents. We do so in the context of China, which is an interesting case. On the one hand, China has over 20 million private firms, some of which are globally leading firms, which suggests considerable entrepreneurial activity. On the other hand, rent-seeking is prevalent, as China not only has a large market in which rents can be extracted but also has the world's largest state-owned enterprises as well as corrupt government officials who control important resources. Thus, *a priori*, it is unclear whether talented Chinese are more or less likely to become entrepreneurs.

We link the universe of 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 created approximately 170,000 firms by 2015) to examine who have become entrepreneurs and how successful their firms are.[1] We supplement the administrative data with a large survey of Chinese college graduates that we conducted during 2010–2015 to study

[1]The median age of firm owners in our matched sample was 33 in 2015, which is comparable to the median age (33.9) for college-educated firm owners in the entire firm registration data. The firm registration data provide self-reported information on whether the owner has a college education. Thus, our data are not particularly biased toward young people.

wage jobs. Our data have several strengths that allow us to address the empirical challenges. First, by observing firm characteristics, we are able to define entrepreneurship and firms more precisely. More recent literature has documented that it is important to consider the definition of firms (as opposed to self-employment) carefully when studying entrepreneurship (e.g., Schoar 2010, Levine and Rubinstein 2017). Second, having measures of firm success, which are normally difficult to observe in other data, enables us to test whether our measure of general talent can be translated into entrepreneurial ability. Third, we can focus on a within-college comparison and, thus, avoid important confounding factors, such as college network and reputation, that affect firm creation.[2]

Our measure of talent is the score on the National College Entrance Exam (known as *gaokao*), which is the criterion for college admission and is popularly believed to determine the course of one's life.[3] China is not unique in this regard, as many other countries also rely on exam scores for college admission (e.g., Chile, South Korea, Turkey, and Vietnam). Nevertheless, it remains unknown whether these exam scores capture general talent. We show that the score is positively associated with both firm success and wages, suggesting that, indeed, it may measure some general ability usable for both entrepreneurial activities and wage jobs. Of course, the score is also associated with socioeconomic status and effort put into the test. These associations, however, should generate relationships opposite to our main findings, as we will discuss.

We find that higher-score individuals are less likely to create firms. 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 and negative association between the score and firm creation: A one-standard-deviation higher score is associated with a 10% lower probability of creating a firm. We focus on within-college analyses for conceptual and empirical reasons. Conceptually, the within-college comparison helps to control for the confounding factors of college reputation and network. Empirically, we find that most of the variation in firm creation comes from within colleges. For instance, although college fixed effects can explain up to 19% of the variation in wages of paid jobs (in our survey data), the effects can account for only 1.2% of the variation in firm creation. In addition, we find that the college fixed effects can explain only 46% of the variation in exam scores, leaving the majority of the variation to occur within colleges.[4]

[2]As highlighted by a survey of approximately a hundred studies on education and entrepreneurship (Van der Sluis, Mirjam Van Praag, and Vijverberg 2008), an important challenge to interpret the education-entrepreneur relationship is to separate ability from education.

[3]For instance, Jia and Li (forthcoming) show that a few points in the exam can make a big difference in terms of access to elite education and obtaining better jobs after graduation.

[4]This is due to the uncertainty in the admission system and political economy of provincial quota assignment. See more discussion in Section 2.

The negative score-firm creation relationship holds even when we focus on the creation of more successful (large) firms, and the magnitudes are similar when using these alternative definitions of firms. We also find that firm creation is positively correlated with observable family background variables (which are generally 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. We further find that the negative score-firm link holds when we examine the relationship by majors and by college quality, implying that the relationship is general.

Does the negative score-entrepreneur relationship support the talent allocation hypothesis? Specifically, the *"talent allocation"* hypothesis assumes that the score reflects a general talent of people that can be used in both entrepreneurial and other sectors, but higher-ability individuals are lured away by other sectors. In principle, however, there could exist two alternative hypotheses, broadly speaking. In the first alternative hypothesis, which we term the *"lower entrepreneurial ability"* hypothesis, the score does not measure a general talent that is useful for entrepreneurs; rather, it is negatively correlated with entrepreneurial ability, which could potentially explain the negative correlation between score and firm creation. For instance, one can assume that those with higher scores lack the skills of jack-of-all-trades (e.g., Lazear 2004) and/or that they obtain higher scores by putting more effort into the test, but their actual ability may be lower than that of their lower-score counterparts. The second alternative hypothesis emphasizes that higher-score individuals possess unfavorable behavioral traits (e.g., more risk averse or less social) for becoming entrepreneurs, which we term the *"personal traits"* hypothesis. Building on earlier work (e.g., Lucas 1978; Evans and Jovanovic 1989; Levine and Rubinstein 2019), we formalize all three possible hypotheses in a simple conceptual framework and empirically test them.

An important test to differentiate the *"talent allocation"* hypothesis from other hypotheses is to examine whether the score is predictive for success in both entrepreneurial and non-entrepreneurial sectors. Exploring the same within-college variations, we first find that the score is positively correlated with entrepreneurial success (starting larger firms and being more likely to expand, enter non-local markets, survive, and be publicly listed), suggesting that the talent, as measured by the score, can be turned into entrepreneurial ability. These results hold when we address potential entry bias by using within-college peer exposure to predict entry. Using survey data, we then find that the score also is correlated with earnings of wage jobs, suggesting that talent, as measured by the score, also can be applied to non-entrepreneurial sectors. In addition, higher-score individuals are more likely to obtain jobs that provide local *Hukou* (i.e., residency rights in a city) and, thus, better access to important local public goods in this context. These results suggest that the score measures some general talent that is valued by both the entrepreneurial and non-entrepreneurial sectors.

We then test, using both survey and administrative data, whether talented people are reluctant to become entrepreneurs because they are attracted away by alternative opportunities, including rent-seeking. By relying on survey data, we find that the state sector, which presumably is more likely to engage in rent-seeking,[5] plays a relatively larger role in attracting talented individuals, even though both the state and private sectors are important opportunities for potential entrepreneurs. We find that, with a one-standard-deviation higher score for an individual, the relative likelihood of working in the state sector is about 30% larger than becoming an entrepreneur and that of working in the private sector is around 18% larger. Using administrative data on firm entry by industry, we document wide variation across industries. For instance, the negative score-entrepreneur correlation in the construction industry is 15 times of that in the IT industry. We find that state penetration in an industry (measured by the share of state investment in total fixed investment) appears to be systematically associated with entrepreneurial reluctance of higher-score individuals: One log point in state penetration is associated with a 59% decrease in the score-entrepreneur relationship across 19 industries, which partly explains the difference between the construction and IT industries.

Our findings do not seem to support the *"lower entrepreneurial ability"* hypothesis. Although we agree that effort and family input matter for scores, our finding that higher-score entrepreneurs are more successful in all performance metrics suggests that the score is indeed a positive measure of entrepreneurial ability. Our findings also do not seem to support the *"personal traits"* hypothesis. For example, if we used this hypothesis to explain the variations by industry in the correlation of score and firm creation, we would have to assume that personal traits vary in the same manner by industry. This would be too strong an assumption. In addition, using survey data we can also directly examine the correlations of scores and other personal characteristics directly. We find that, within-college, higher-score individuals have a higher GPA and are more likely to receive academic awards and to become a Chinese Communist Party (CCP) member. We do not, however, find strong correlations between scores and risk attitudes or participation in social activities. Overall, these results suggest that scores are not correlated with these behavioral traits in a way that is particularly unfavorable for becoming an entrepreneur.

Our findings provide evidence for the theoretical literature that talented individuals can join both the productive and unproductive sectors, depending on the reward structure (e.g., Baumol 1990; Murphy, Shleifer and Vishny 1991, 1993; Acemoglu 1995; Baumol, Litan and Schramm 2007). Few

[5]This is not to deny that it is useful to attract some talented individuals to work for the state. In fact, talented bureaucrats are an important source of the Chinese state capacity. However, when a large number of talented individuals choose to work in the bureaucracy, the government sector can generate brain drain. See more discussion in Section 5.2.

studies have directly tested hypotheses generated from these theoretical insights. Our data allow us to examine how a talent measure correlates with success in different sectors and, thus, support a key assumption – talent can be general – in this theoretical literature. We also find evidence that the rent-seeking sector can lure away talent, as emphasized by the theory. Our results appear to be in contrast to two studies that use data from the United States: Levine and Rubinstein (2017) find that, in the United States, talented people (with higher scores in Armed Forces Qualifications Test) are more likely to own incorporated businesses;[6] Shu (2016) finds that higher-GPA graduates from M.I.T. are more attracted to science and engineering than to the finance sector. This contrast could likely be a result of institutional differences between the two countries; e.g., rent-seeking is more prevalent in China. We indeed find that talented people are less likely to become entrepreneurs in industries for which the state plays a more dominant role. In the Chinese context, it is recognized that the dominance of the state sector potentially hinders economic development (e.g., Huang 2008). Brandt, Kambourov, and Storesletten (2020) construct a quantitative model that highlights the importance of entry barriers to explain productivity differences across regions and show that these barriers are related to the size of the state sector. Our study provides individual-level support for these macro-regional approaches.[7]

While we cannot give a fair review of the extensive literature on entrepreneurship,[8] our contribution is to better understand the relationship between individual talent and firm creation, separating the impact of talent from that of education. Our results on the positive association between individual scores and firm success also are consistent with a growing literature that emphasizes the role of entrepreneurial human capital in determining firm performance (Bertrand and Schoar 2003; Bloom and Van Reenen 2007; Gennaioli et al. 2013; Queiró 2018; see an excellent survey by Syverson (2011)).

Finally, our study joins the recent empirical literature that tries to understand talent misallocation

[6]In addition, Levine and Rubinstein (2017) find that, among the individuals with high scores from the Armed Forces Qualifications Test, those who exhibit more illicit behavior are even more likely to own incorporated businesses. In our context, it is also possible that there exist interaction effects between the score and non-cognitive skills, which is not our focus. Our main interest is to test the talent allocation theory, which concerns the relative returns to the absolute advantage, measured by the score, across sectors.

[7]We do not focus on regional variation in this study because the individuals whom we study all have college degrees and are mobile, most of whom chose to work in big cities rather than in their hometowns. It is thus difficult to associate regional state penetration with their career choices. Empirically, we find that the negative score-firm relationship holds across provinces.

[8]See Parker (2018) for an extended discussion of the literature, in which he reviews individual determinants and macro factors that influence entrepreneurship. China, with its economic growth and increasing market size in recent decades, offers fertile research ground for this topic. See also recent works on how community networks shape private entrepreneurship (e.g., Dai et al. 2020).

in different settings (e.g., Hsieh et al. (2019) on racial discrimination in the U.S. and Ashraf et al. (2021) on gender norms across countries). More broadly, our study is also related to a burgeoning literature using administrative data from developed countries to study the social and economic background of politicians (Dal Bó et al. 2017) and inventors (Aghion et al. 2017; Bell et al. 2019). Our findings suggest an important role of the state in shaping talent allocation, which could be a useful perspective to understanding variations in talent allocation across countries.

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*), administered 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 (exam subjects are Chinese, English, advanced math, and sciences) and social sciences (exam subjects are Chinese, English, basic math, and social sciences) tracks. At the end of the third year (normally in June), students take the National College Entrance Exam in one track. Because the exams and admissions are administered by each province every year, the exam scores are comparable only for students from the same province and year, as well as the same track (social or natural sciences).

We employ administrative data that cover the entire universe of all participants in *gaokao* from 1999 to 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 are the data that we use. The data provide detailed information on student exam performance, including total score, subject scores, name of college and major (if admitted) as well as student biographical information, such as gender, *Hukou*, birth year, and high school name.[9]

To track firm creation and performance, we use administrative data of all firms ever registered in China by February 2015.[10] These data include 28 million firms, 11 million of which had been

[9] 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 that, in 2011, China's college dropout rate was 3%, whereas the Ministry of Education reported that the dropout rate that year was 0.75%.

[10] For small businesses, 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. Notably, however, remaining self-employed without registering a firm may have implications that contribute to tax evasion. We consider both small and big firms, for which registration is

de-registered by 2015. The administrative registration records provide information on the owners (known as shareholders), which can be individuals or firms. We call individual owners firm creators or entrepreneurs. We also know whether a firm invested in another firm as an owner, which we define as an "expansion". The data also include 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.

The college entrance exam and admission data are linked to the firm owner data by an encrypted national identity number in both datasets. Of the 14.2 million records, we succeed in linking 9.5 million records,[11] with 9.1 million unique individuals. The extra 400,000 records are repeat exam takers, for whom we focus on the first-time score because we are interested in score as a proxy for ability. We examine how missing data and the existence of repeat exam takers affect our results. In addition, a small share of students received extra points due to non-exam experience such as being ethnic minority, which we exclude from our score measure. 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 2010–2015

We supplement the administrative data with our own college graduate survey data, which include information on wages. We conducted large-scale surveys of college graduates during the graduation months (May and June) between 2010–2015, which cover approximately 30,000 students from 90 colleges, around 14,800 of whom reported detailed information on their first jobs. We designed the surveys to evaluate the elite college premium and intentionally asked about *gaokao* scores (see Jia and Li (forthcoming) for a detailed description of these surveys and how the first job is important for future jobs). Moreover, the surveys include information on student performance and behaviors in college. Although these surveys cover only a limited number of colleges, this is not a major issue for us, as our main interest is a comparison of students who attended the same colleges. The individuals in the survey were born later than those in our administrative data, which can be considered an "out-of-sample" test of our findings.

required.

[11] Matching of the two datasets is not perfect due to missing (3.9 million) or invalid identifying numbers (0.8 million). The missing of identifying numbers usually occurs at a province-year level. We further restrict our analyses to province-years with few missing numbers as a robustness check.

2.3 Variables and Summary Statistics

Firm Variables We present firm-level variables in Panels A and B 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% of the total.

In light of the recent literature that emphasizes 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,[12] which is the maximum liability that 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 observe only the most current number in 2015. If a firm exited before 2015, its registered capital size refers to the most recent information before the exit. Using registered capital as a measure of firm size, we find that 25th and 75th percentiles are RMB 200,000 (1 USD = 7 RMB) and 2 million. Of the college graduates, 1.6% had established firms over RMB 2 million and only 0.4% above RMB 15 million. We call firms with registered capital over RMB 2 million medium-sized firms and those above RMB 15 million large firms, which are relative concepts for trackability.

In addition to firm size, we use a few alternative variables to measure firm success. The first is whether the firm was registered out of one's home province. It is 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, due to local protection, it is not easy to establish a firm beyond one's home province. The second variable is expansion, which is defined as one firm's investing in another firm as an owner. In our sample, 5% of firms expanded this way. The third variable is whether a firm became publicly listed. This measure can be considered 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 three years, but only 60% survived beyond 10 years. Notably, each measure has its own limitations. For instance, the size of the registered capital partly reflects credit access; deregistration is only a noisy measure of exit. Altogether, however, the five variables allow us to achieve a good understanding of entrepreneurial success.

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

It is useful to note that most of the variation in firm creation comes from within colleges. College fixed effects account for only 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. This suggests that the choice of entering the entrepreneurial sector varies greatly at the individual level.

Exam Scores Although exam scores are positively correlated with college quality, there exists considerable variation in scores within colleges. To see this variation, 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, etc.). On average, better colleges accepted higher-score students. College fixed effects, however, can explain only less than half (46%) of the variation in exam scores. Exam scores vary greatly within colleges, evidenced by the overlapping of scores across different groups of colleges in Figure 1(a). To measure the variation, we report in Panel C of Table 1 the mean and standard deviation of the exam scores, which are 436.9 and 101.4 respectively. Once we control for province-year-tracks fixed effects, the standard deviation drops to 88.9. It goes down to 68.4 after we also control for the college fixed effects, and 66.7 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 is based on a few institutional reasons. First, college application and admission is a highly uncertain process. In our study period, students in most provinces applied for colleges before they knew their exam scores. Each exam taker needed to indicate college preference, via a 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 nearly useless). As a result, the matching between score and college was far from being ideal.[13] Second, scores for each college vary greatly across provinces due to the uneven distribution of admission quotas. Each province is assigned a quota for each college by the central government. Due to political and historical considerations, major metropolitans, such as Beijing and 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,

[13] Such uncertainty was marginally mitigated only in recent years, when the admission system was reformed to become a parallel system that, thanks to computer technology, allows students to apply to a few more colleges.

such as economics, finance, law, and STEM, are normally higher than those for humanities. We consider the role of majors in our analyses. Thus, the first two institutional reasons are the main driver for the within-college variation in exam scores.

Personal Background We report summary statistics of personal attributes in Panel C of Table 1. As seen in the table, 54% of the college students are male, slightly higher than the male share in the population (51.3% in 2001), and 53% of the students have urban *Hukou*, much higher than the urban share in the population (37% in 2001), which is consistent with the fact that fewer rural students are able to attend high school. The median birth year of firm owners in our sample is 1982, or age 33 in 2015, close to the median age of college-educated firm owners in the entire firm registration data (33.9).

We run regressions to examine how the *gaokao* score varies with owner characteristics. As shown in Appendix Table A, although males do better across colleges (Column (1)), females and males do not differ in their *gaokao* score within colleges (Column (2)). Rural students have higher scores, both across and within colleges, most likely because only better students from rural areas can attend high school and take the exam.

We use two variables to proxy family socioeconomic status. The first one is high school quality, as wealthier families are more likely to be able to afford elite high schools. For instance, Ye (2015) shows that the parents of those in better high schools receive more years of education and are better paid. We measure high school quality using the share of students in each high school who are admitted by the top-100 colleges in China.[14] As reported in Table 1, the bottom 25% and top 25% high schools have 4.8% and 19.4%, respectively, of their students admitted by the top-100 colleges, indicating a wide variation across high schools. In addition, we use the income per capita of 2001 in one's birth county as a measure of individual income. Not surprisingly, students from better high schools and wealthier counties tend to have higher scores, both across and within colleges, as shown in Table A.

3 Scores and Firm Creation

In this section, we present the findings for our test of whether talented people are more likely to become entrepreneurs. Empirically, we examine the sign and size of the correlation between the *gaokao* score and firm creation at the individual level and how it varies by major and college tier.

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

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 use three definitions of firms: any size, medium-sized, and large firms. We isolate province-year-track fixed effects so that the scores are comparable.

Descriptive results suggest that the correlation between firm creation and exam score is generally negative, but even more so when we control for college fixed effects. As shown in Figure 2(a), the score-firm creation 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(b) and (d).[15]

Baseline Estimates To estimate the link between scores 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 $Firm_{i,pyt,c}$ is a dummy variable that indicates whether individual i of province-year-track (pyt) in college (c) created a firm. Again, the dependent variables are the creation of firms by size. The key independent variable of interest is the exam score ($Score_{i,pyt,c}$). X_i indicates one's personal characteristics, which include gender, *Hukou* (rural vs. urban), high school quality, birth county's GDP per capita, and age fixed effects. Although we do not have measures such as parental income, it is reasonable to assume that those from better high schools and wealthier counties have higher socioeconomic status. Because $Score_{i,pyt,c}$ is comparable only 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 that are clustered at the college level. We also control for the major fixed effects in X_i in some specifications.

The estimation results reported in Table 2 show that firm creation is negatively correlated with exam score. To make interpretations easier, we report the coefficient for one-standard-deviation difference in the exam score (i.e., 68 points in a college). As shown in Column (1), within a college, a one-standard-deviation higher exam score decreases the probability of creating any firm by 0.76 percentage points, or over 10% of the mean (7.21). The estimates remain stable when we control for all personal characteristics, as seen in Column (2), and add 12 major fixed effects, as seen in Column (3). If we define firms as those with registered capital of no less than RMB 2 million or 15

[15]In addition, Appendix Figure B.1 shows that the negative association between the score and firm creation within colleges holds after controlling for 12 major fixed effects.

million, the magnitudes of the estimates are actually very similar when they are benchmarked to the mean, still around 10% of the mean, as reported in Columns (4)–(9).

In addition, we find that personal attributes matter for firm creation. The gender difference is large: Females are only half as likely to create a firm as are males, likely because females 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 an advantageous social and economic background also have an advantage in creating firms.

Major and College Rank Theoretically, the relationship between scores and firm creation may differ by college major and college quality due to different opportunity costs. In the above analysis, we control for major and college dummies and find that the score-firm creation correlation holds. We now go a step further to allow the effect of scores on firm creation to vary across major and college quality. To track our analysis, we categorize the 12 majors into three groups: STEM, economics-finance-law, and humanities. 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-law majors account for 27% of the students and humanity majors, for the remaining 21%. We also divide colleges into three groups by quality: the top 10, those ranked 11–100, and the rest.

The relationship between the exam score and firm creation does not change dramatically across major and college quality. As shown in Table 3, students who are economics-finance-law majors have a higher mean probability of firm creation, consistent with the conjecture that the choice of these majors indicates a preference for entrepreneurship. Most importantly, the negative relationship between scores and firm creation holds within each major. For instance, for creating any firm (Columns (1)-(3)), a one-standard-deviation higher exam score within a college decreases the probability of creating firms by 8.8% for STEM majors, 7.6% for economics-finance-law majors, and 12% for humanities majors. Similarly, as shown in Table 4, those who go to better colleges are more likely to start a firm, but the negative association between scores and firm creation holds for each college-quality group. Thus, the negative score-entrepreneur relationship is general in terms of different majors and colleges.

Potential Data Issues We check the sensitivity of our analysis by considering a number of data issues. First, to check whether missing data due to matching is an issue, we restrict our analysis to province-years with missing probabilities lower than 5% and 1%. As reported in Panels A and B of Appendix Table B.2, the results remain similar, suggesting that the missing data may not have been a critical issue in our case.

Second, we exclude repeat exam takers from our analysis to examine whether including them caused a problem. Again, we obtain results very similar to our baseline estimates (reported in Appendix Table B.3).

Third, as indicated in Figure 1, there is 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 conduct an analysis in which we remove 10% of observations from both tails. The estimated negative correlations become even stronger, as shown in Appendix Table B.4.

Finally, our firm data include family firms. The concern is that college graduates may come to these firms through succession rather than through creation. The data suggest that the share of family succession in our sample is likely to be small, as over 98.2% of the firms were established after the exam year of the students. We can also gauge the extent of family succession by examining the age difference between shareholders. In this exercise, we exclude potential family firms, i.e., firms that were established before the owner took the exam and firms with an age difference between our college-educated owner and the eldest shareholder in the firm of more than 20 years. Excluding these firms, we again find that a one-standard-deviation higher exam score is associated with a 10% lower probability of firm creation (Appendix Table B.5).

4 A Conceptual Framework

In this section, we attempt to explain our finding that higher-score individuals are less likely to become entrepreneurs. We consider three competing hypotheses, revolving around the sign of the correlation between scores and entrepreneurial ability, the *"talent allocation"* hypothesis, the *"lower entrepreneurial ability"* hypothesis, and the *"personal traits"* hypothesis.

To better understand how these three hypotheses work to generate the negative correlation between the score and firm creation, we present a stylized model, building on earlier literature (e.g., Lucas 1978; Evans and Jovanovic 1989; Levine and Rubinstein 2019). Our model considers all three hypotheses, and we derive testable predictions which we then bring to the data in the next section.

Setup An individual with score S (we ignore the subscript i to simplify notations) chooses to either establish a firm or become a worker. The score 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

$$A_E = e^{aS} \text{ and } A_W = e^{bS}. \tag{2}$$

a and b can have any sign, reflecting that the correlations between the score, entrepreneurial, and wage abilities can be positive, negative or zero.

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

$$Y = A_E^{\beta} K^{\alpha},$$

where β captures the return to entrepreneurial ability.

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

$$R = A_E^{\beta} K^{\alpha} - rK,$$

where r is the price of capital. Solving this, we get the optimal amount of capital

$$K^* = \left(\frac{\alpha A_E^{\beta}}{r} \right)^{\frac{1}{1-\alpha}}. \tag{3}$$

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

$$\begin{aligned} \ln R^* &= \tfrac{1}{1-\alpha} \ln A_E^{\beta} - \tfrac{\alpha}{1-\alpha} \ln r + \ln \Gamma \\ &= \tfrac{1}{1-\alpha} aS\beta - \tfrac{\alpha}{1-\alpha} \ln r + \ln \Gamma, \end{aligned} \tag{4}$$

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

$$W = A_W^{\mu},$$

where μ captures the return to wage ability.

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

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

and taking a log we get

$$\begin{aligned} \ln V &= \mu \ln A_W + \delta(S) \\ &= bS\mu + \delta(S). \end{aligned} \tag{5}$$

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 derive

$$\frac{1}{1-\alpha} aS\beta - \frac{\alpha}{1-\alpha} \ln r + \ln \Gamma \geq bS\mu + \delta(S) + \varepsilon. \tag{6}$$

Thus, the probability of creating a firm becomes

$$\mathbb{P}(Firm) = \mathbb{G}\Big\{ \left[\frac{1}{1-\alpha} a\beta - b\mu \right] S - \delta(S) - \frac{\alpha}{1-\alpha} \ln r + \ln \Gamma \Big\}. \tag{7}$$

From equation (7), we can see a negative link between the score and firm creation (dP/dS) is equivalent to the following inequality,

$$\frac{1}{1-\alpha} a\beta - b\mu - \delta_s(S) < 0. \tag{8}$$

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

(1) The first hypothesis is the *"talent allocation"* hypothesis, which assumes $a > 0$ and $b > 0$, and $\frac{1}{1-\alpha} a\beta < b\mu$. Here, the score is positively correlated with both the entrepreneurial ability and wage ability, but the return of the wage ability to score rises faster with the score than that of the entrepreneurial ability, leading to a negative relationship between the score and firm creation. In other words, talents are allocated to the non-entrepreneurial sector because of the high opportunity cost.

(2) The second or the *"lower entrepreneurial ability"* hypothesis assumes a negative correlation between the score and entrepreneurial ability, which refers to $a < 0$. There can be different reasons for why higher-score individuals have lower entrepreneurial ability. For instance, those with higher scores may lack the skills of jack-of-all-trades. Another possibility is that individuals achieve higher scores mainly via effort we cannot observe. If one could isolate effort from the scores, one might find that the higher-score individuals have a lower entrepreneurial ability.

(3) The third hypothesis, which we term the *"personal traits"* hypothesis, does not assume any correlation between the score and entrepreneurial ability. Instead, the score reflects other personal traits that are unfavorable to become an entrepreneur, or $\delta_s(S) > 0$. For example, those scoring higher might be more risk averse, or they are from family backgrounds that limit their chances of starting firms.

Empirical Tests To test the *"talent allocation"* hypothesis versus the *"negative entrepreneurial ability"* hypothesis, we recover the sign of a by studying the relationship between the score and firm outcomes. In particular, we know

$$\ln K^* = \frac{1}{1-\alpha}\left[aS\beta + \ln \alpha - \ln r\right]. \tag{9}$$

The estimated sign of dK/dS is the same as a, given that $\beta > 0$. In other words, using K as a performance measure, the effect of the score on performance has the same sign as estimated a. This logic applies to other firm success measures.

Another test of the *"talent allocation"* hypothesis is regarding the sign of b. We can estimate the sign of b by estimating a Mincer log wage equation, having the score (S) as an explanatory variable,

$$\ln W = \mu b S. \tag{10}$$

In addition, the *"talent allocation"* hypothesis also predicts $\frac{1}{1-\alpha}a\beta < b\mu$. Although we do not have income data to directly estimate each parameter, it is useful to check whether this inequality is related to support the importance of relative returns (β and μ). To do so, we examine the relationship between the score and firm creation by industries. Under the assumption that the mapping precisions between the score and ability (a and b) are similar across industries, a wide variation in the relationships between the score and firm creation across industries supports the importance of relative returns (β and μ).

Finally, we test the *"personal traits"* hypothesis by directly examining the relationship between the score and personal traits, which will help us to get a sense of the sign of $\delta_s(S)$. We will link the score to variables such as students' in-college academic performance, CCP membership, participation in social activities in college and risk attitudes.

5 Evidence for the Talent Allocation Theory

In this section, we present the results when we examine whether the negative score-entrepreneur relationship supports the talent allocation theory: Talented individuals are attracted away from entrepreneurship by alternative opportunities. We first test an important assumption of the theory: There is some general talent of people that can be used by both entrepreneurial and other sectors (i.e., $a > 0$ and $b > 0$). We show that the *gaokao* score, presumably a measure of general talent, is correlated with success in both entrepreneurial activities and wage jobs. We then show that rent-seeking might indeed be behind the entrepreneurial reluctance by examining how the size of the state sector affects the chance of becoming an entrepreneur.

5.1 Scores and Firm Success

We first examine whether the score measures the talent that is useful for entrepreneurial success. If higher-score individuals have higher entrepreneurial ability, we should expect to see that their firms are more successful, as illustrated by equation (10).

Empirical Strategies Empirically, we examine the relationship between scores and firm success using the following specification:

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

where $y_{f,i,pyt,c}$ refers to different success measures for firm f created by individual i. Similar to our analyses above, we remove the impact of colleges by including college fixed effects.

One concern, however, is selection bias. The firm outcomes we observe are conditional on entry, which may be correlated with unobservable variables that also can affect success. For instance, there might be individual-level shocks in the firm creation decision that are correlated with firm success. If such shocks vary systematically with the score, 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 employ 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 n refers to the number of students in a college-cohort, and $EntrepreneurProp_{j,prov}$ 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 about the reflection problem.

The differences between the estimates in Columns (1) and (2) of Table 5 illustrate the advantage of this IV. Without controlling for college fixed effects, we find that the score is strongly correlated with the peer entrepreneurial propensity, peer-origin$_c$ (Column (1)). In contrast, the correlation is close to zero once we control for the college fixed effects (Column (2)), suggesting that the within-college exposure is close to being random. This is also true for other personal characteristics, as reported in Appendix Table C.1. Within a college, peer-origin$_{c,t}$ is a strong predictor of firm creation (Column (3)). In terms of the marginal effect, a one-standard-deviation increase in $peer_{c,t}$ is associated with a 2.4-percentage-point higher probability of firm creation, or 25% of the mean.

Results for Firm Success We use a few measures for firm success and find that they are all positively correlated with the score. Our first measure is firm size.[16] Employing log capital size demeaned by industrial mean as the outcome (Columns (4) of Table 5), we find that a one-standard-deviation increase in within-college scores is associated with a 0.6% larger registered capital. The small and imprecise estimate partly reflects the fact that this is a noisy measure of success.[17] The positive links between scores and firm success are stronger when we use alternative measures of firm success, including whether a firm is located outside one's home province, firm expansion, and becoming publicly listed. For these three measures, we find positive and sizable correlations between exam scores and firm success (Columns (5)-(7)): A one-standard-deviation higher within-college exam score increases the probability of investing out of one's home province and of expansion by about 10% (i.e., 0.036/0.394 and 0.005/0.049). In regard to the rare event of becoming listed, the estimate is less precise, but the magnitude of the increase is even larger, about 19% (0.069/0.359) of the mean.

As a comparison, we also report simple OLS estimates (conditional on entry) without correcting for selection bias. The results reported in Panel B of Table 5 and Appendix Figure C.2 show that the OLS estimates are similar to the Heckman estimates, suggesting that selection bias is not quantitatively important, conditional on a valid IV.

We also examine the quality of the IV, i.e., whether peer-origin$_{c,t}$ also correlates with other covariates of firm success. We should note that the finding of no correlation between peer-origin$_{c,t}$ and observables (scores and other personal characteristics) in Column (2) of Table 5 suggests that

[16] Because different industries have different regulations for the minimal entry capital size, we demean a firm's log capital size by the industrial mean.

[17] As noted above, registered capital size is also related to credit access.

peer-origin$_{c,t}$ may not be highly correlated with unobservables. As another test on the exclusion restrictions, we create two IVs by the origin of peers, those from the same region versus those from different regions (region means northern or southern China). If unobservables are important, we should expect to see a larger two-step estimate when using peer exposure from the same region as the IV, as 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 is not quantitatively important.[18]

As another way to measure success (or failure), 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%.[19] As shown in Appendix Figure C.4, firms founded by those with higher scores are more likely to survive, and the advantage increases over time. In a Cox regression, the estimated hazard ratio with respect to the increase of one-standard-deviation within-college exam scores is around 0.975, with a standard error of 0.006. The smaller-than-1 ratio is consistent with the illustration that shows that firms created by higher-score students are less likely to exit.

To summarize, we find that score is positively correlated with entrepreneurial success, and the importance of scores is more apparent when we use stricter definitions of success. These results suggest that score indeed measures some general talent that can be turned into entrepreneurial ability.

5.2 Scores and Wage Jobs

First-job Wages and Benefits Employing our college graduate survey data on first-job wages, we estimate the relationship between scores and log wages, using a within-college specification similar to equation (1).[20] Although the first-job wages are typically compressed, we find that score is, indeed, positively correlated with wages. As shown in Columns (1)–(2) of Table 6), a one-standard-deviation increase in the score is associated with 2.6–2.9% higher first-job wages. Further, males, urban individuals, and those from better high schools also have higher wages than their counterparts, confirming that they enjoy advantages in the job market.

Another important measure of job benefit is whether a job provides local *Hukou*, which

[18] Another, similar exercise is to divide peers by their success, assuming that less-successful peers are less likely to help their own firms. Again, we find similar patterns when using more- or less-successful peers to predict entry.

[19] Specifically, we first obtain the within-college score distribution by isolating province-year-track fixed effects and year fixed effect. Then, we divide the residual scores into five quartiles.

[20] All the results using the survey data have considered the sampling weight, i.e., the regressions are weighted by the inverse of sampling weight in our surveys.

determines whether the worker and family can access local public goods such as education and health care. Using the same within-college specification, we find that higher-score individuals are more likely to get jobs that provide local *Hukou* (Column (3)). This finding holds when we exclude those who had been local before going to college (Column (4)), who thus do not need a *Hukou* provided by employers. Thus, the score is rewarded by wage jobs, in terms of both wages and important benefits associated with jobs.

Lure of the State Sector Our finding that the score is important for both entrepreneurial success and earning higher wages suggests that the score may have indeed measured some general talent, and this talent is useful for both sectors. Here, we further test whether talented individuals are lured away partly due to the opportunity for rent seeking, emphasized by the theoretical literature. One particular sector related to rent-seeking in China is the state sector, which includes governments and state-owned enterprises. Corruption and rent-seeking are prevalent in the government sector, and China was ranked 83rd in 2015, according to the Corruption Perception Index. As to state-owned enterprises, it has been well documented that their productivity, especially capital productivity, is significantly lower than that of their private counterparts, partly due to their advantages in access to capital (e.g., Hsieh and Song 2015). It also has been conjectured that good students tend to be attracted to the state sector. For instance, Cook et al. (2020) provide descriptive evidence on the ideal jobs that college graduates hope to obtain in China, for which jobs in the governments and state-owned enterprises are ranked the highest. In our College Graduate Survey data, over 62.5% of the college graduates chose governments and state-owned enterprises as their ideal jobs.

We further examine, using survey data, the choices of working in the state sector or in a wage job in the private sector versus becoming an entrepreneur. In our survey, 43.4% of the graduates worked in the state sector, 50.4% worked in private firms, and 6.1% became entrepreneurs.[21] We find that the state sector plays a larger role as an alternative opportunity than does the private sector. In Columns (5)-(8) of Table 6, we report multinomial logit regression results on how the score is associated with the relative risk of working in the state and private sectors versus becoming an entrepreneur. The relative risk of working in the state sector for individuals with one-standard-deviation higher score is 1.28–1.32 times higher, and it is significantly different from 1 (see the p-values in Columns (5) and (7)). The relative risk for higher-score individuals to work in the private sector is around 1.18, but it is not significantly different from 1 (Columns (6) and (8)).

[21] Unlike the administrative data that focus on firms, the survey did not differentiate between firm owners and the self-employed.

5.3 Variation by Industry

Motivated by the above finding that the state sector plays a larger role in attracting higher-score individuals, we want to know whether the negative link between score and firm creation varies systematically with the size of the state sector across industry. We estimate equation (1) by industry and plot the 19 industry-specific estimates in Figure 3(a). Here, the dependent variable is a dummy that indicates whether an individual created a firm in a certain industry (i.e., the variable is zero if he or she does not create a 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: On one end of the spectrum, with an increase of the score by one standard deviation, the probability of entering the construction, mining, public management, culture, real estate, and residential sectors decreases by 25–35%; on the other end, with an increase of the score by one standard deviation, the probability of entering the restaurant and hotel, science and technology service, and IT sectors decreases by less than 5%.

Our main interest is to determine whether the heterogeneities across industries reflect the degree of state penetration into certain industries. To measure the importance of the state in an industry, we use the share of state fixed investment (relative to total investment) in 2010, which we term *state penetration*.[22] The variation across industries, however, partly reflects the importance of human capital across industries. To take this into account, we control for the average schooling years of employees in each industry.[23] After controlling for schooling, we plot the residual relationship between the share of state fixed investment and the estimated correlation between the score and firm creation by industry.[24] As shown in Figure 3(b), one log point in state penetration is associated with a -0.059 decrease in the impact, or around 59% of the mean. Thus, these results imply that higher-score individuals are more likely to avoid creating firms in industries with stronger state penetration, which further supports the talent allocation theory.

6 Alternative Interpretations

Can the *"lower entrepreneurial ability"* and the *"personal traits"* hypotheses explain our findings? As shown in the conceptual framework, theoretically, our main story and these alternative interpretations

[22]The fixed investment comes from the China statistical yearbook in 2011. Because the size of the state sector is fairly stable within an industry, the pattern we show is robust to using information from alternative years.

[23]We calculate the schooling by industry from the 2010 census.

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

could all work together to drive the negative correlation between the score and firm creation. Below, we discuss the relevance of the two broad alternative interpretations.

6.1 Lower Entrepreneurial Ability

The conjecture that higher-score individuals have lower entrepreneurial ability is based on multiple reasons. For instance, those with higher scores lack the skills of jack-of-all-trades; they obtain higher scores by putting more effort into the test, but their actual ability could be lower than their lower-score counterparts; or they get higher scores because they come from better socioeconomic background and have more resources to help them prepare for the test. Although we agree that effort and family input can affect scores, our results above on firm success suggest that higher-score individuals actually have higher entrepreneur ability (as well as wage ability). Therefore, our data do not support the lower-entrepreneurial-ability interpretation.

6.2 Personal Traits

Is it possible that our finding is driven by the fact that higher-score individuals in a college possess unfavorable traits for becoming entrepreneurs? For instance, they might be more risk averse or less social, which consequently affect firm creation. First of all, this interpretation is difficult to reconcile with our finding that the score-firm creation correlation varies with industry, unless we can make a strong assumption that the relationship between scores and these traits also vary by industry in the same manner. Nevertheless, using our survey data, we can examine directly the correlations between the score and personal traits, including in-college academic performance, CCP membership, participation in social activities in college, and risk attitudes.

Not surprisingly, we find that higher-score individuals have a higher GPA and are more likely to obtain academic awards than lower-score college peers (Columns (1)-(2) of Table 7). Higher-score students also are more likely to become a CCP member (Column (3)), consistent with our earlier finding that they are also more likely to enter the state sector, in which CCP membership is often valued. We do not find much of a relationship between scores and participation in social activities, whether we measure by having a position in the college student union or in any social organization (Columns (4)-(5) of Table 7). Overall, we do not find supporting evidence for a negative relationship between scores and social skills.

Finally, we focus on risk attitudes, an important determinant of entrepreneurship. Conceptually, this risk attitudes hypothesis should be more relevant for individuals with scores at the tails, because those at the margin might get into a college due to their risk-taking behavior. Our finding, however, becomes even stronger if we exclude those at the tails (recall Appendix Table B.4). In addition, we have direct measures of risk attitudes from the 2011 wave of the College Graduates Survey. We asked two risk attitude questions. The first was, "Do you prefer to obtain RMB 1,000 with certainty or play a lottery to get between 0 and 2,000 with an equal chance?" We assume that one is more risk averse if he or she rejected the lottery option. The second question was, "Do you agree with the statement that ensuring certain returns on investment is more important than is taking more risks to gain higher returns?" Risk aversion is a dummy variable that takes the value of 1 if a student chooses to agree or strongly agree with this statement. We do not find strong correlations between scores and risk attitudes, as shown in Columns (6)-(7). Thus, the relationship between scores and risk attitudes is not obvious.

These results suggest that scores are positively correlated with GPA, the probability of getting academic awards, and the probability of becoming a CCP member, whereas their correlations with social activity participation and risk attitudes are not clear. Thus, it seems difficult to argue that higher-score individuals exhibit certain behavioral traits that hinder entrepreneurship. In addition, if we assume that higher-score individuals possess traits unfavorable for entrepreneurship, this assumption would contradict the finding that higher-score individuals are more successful entrepreneurs. That said, our findings do not detract from the importance of risk aversion and other personal traits for entrepreneurship. In fact, some of our findings suggest that personal traits play an important role in firm creation. For example, we find that females, 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 the lure of alternative opportunities. In fact, they have lower wages, as shown in Table 6. These facts suggest that risk aversion and other personal traits matter for entrepreneurship.

7 Conclusion

The theoretical literature on talent allocation has long provided insights into understanding the selection into entrepreneurship versus rent-seeking, which has important implications on economic development. Due to the challenges of measurement, however, few empirical studies have directly investigated the hypotheses generated from these theoretical insights. By leveraging administrative and survey data and focusing on arguably the most important talent measure in Chinese society, we find evidence supporting this literature. In particular, controlling for colleges (and thus their influences on signaling and network), we document that, individuals with higher college entrance exam scores are *less likely* to create firms. This negative relationship, however, is unlikely because higher-score individuals have lower entrepreneurial ability. On the contrary, firms created by higher-score individuals are more successful than those created by their lower-score peers. The opposite patterns for *ex ante* firm creation and *ex post* firm success are more consistent with the interpretation that scores are positively correlated with entrepreneurial ability but that higher-score students are attracted away by non-entrepreneurial sectors, particularly the state sector, evidenced by wage-job sectors and variation across industries in firm creation.

Can we apply our findings to other countries? In principle, one can link individual-level SAT (or other exam) scores to firm creation and firm success in the United States or other countries,[25], however, we are not aware of any such studies. In light of our findings on the importance of the state, however, we conjecture that there may be wide variation across countries, depending on the role of the state in the economy. We thus hope that our study provides new avenues for future research that allow scholars to compare talent allocation across countries.

References

[1] Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen (2017), "The Social Origins of Inventors," NBER Working Paper 24110.

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

[3] Acemoglu Acemoglu, Daron (1995), "Reward Structures and the Allocation of Talent," *European Economic Review,* 39(1): 17–33.

[4] Baumol, William J. (1990), "Entrepreneurship: Productive, Unproductive, and Destructive," *Journal of Political Economy,* 98(5): 893–921

[5] 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.

[6] Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2019), "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *Quarterly Journal of Economics,* 134(2): 647–713.

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

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

[9] Brandt, Loren, Gueorgui Kambourov, and Kjetil Storesletten (2020), "Barriers to Entry and Regional Economic Growth in China," Working Paper.

[25] There are studies that examine SAT scores and mutual fund performance (e.g., Chevalier and Ellison 1999). Such a measure, however, refers to average SAT scores for colleges rather than individual-level scores and, thus,likely combines both college and individual effects.

[10] 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.

[11] Cook, Sinclair, Hongbin Li, Lingsheng Meng, and Yanyan Xiong (2020), "Brain Drain to the State Sector: Job Preferences and Outcomes for China's College Graduates," Working Paper.

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

[13] Dai, Ruochen, Dilip Mookherjee, Kaivan Munshi, and Xiaobo Zhang (2020), "Community Networks and the Growth of Private Enterprise in China," Working Paper.

[14] Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne (2017), "Who Becomes a Politician?," *Quarterly Journal of Economics,* 132(4): 1877–1914.

[15] 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.

[16] 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.

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

[18] Hsieh, Chang-Tai, and Zheng Michael Song (2015), "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China," *Brookings Papers on Economic Activity,* Spring: 295-346.

[19] Huang, Yasheng (2008), *Capitalism with Chinese Characteristics: Entrepreneurship and the State.* Cambridge University Press.

[20] Jia, Ruixue and Hongbin Li, "Just Above the Exam Cutoff Score: Elite College Admission and Wages in China," *Journal of Public Economics,* forthcoming.

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

[22] 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.

[23] Levine, Ross, and Yona Rubinstein (2019), "Selection into Entrepreneurship and Self-Employment," NBER Working Paper 25350.

[24] Lucas Jr, Robert (1978), "On the Size Distribution of Business Firms," *Bell Journal of Economics*: 508-523.

[25] 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.

[26] Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1993), "Why Is Rent-seeking So Costly to Growth?" *American Economic Review,* 83(2): 409–414.

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

[28] Queiró , Francisco (2018), "Entrepreneurial Human Capital and Firm Dynamics," Working Paper.

[29] Schoar, Antoinette (2010), "The Divide between Subsistence and Transformational Entrepreneurship," *Innovation Policy and the Economy*, 10(1): 57-81.

[30] Shu, Pian (2016), "Innovating in Science and Engineering or 'Cashing In' on Wall Street? Evidence on Elite STEM Talent," Working paper.

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

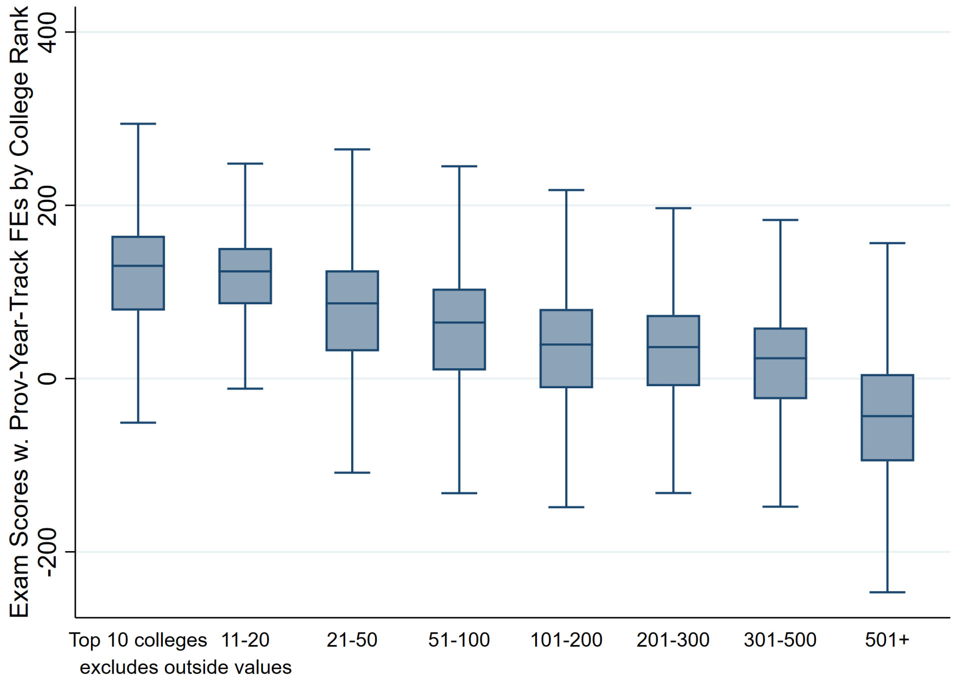
[32] Van Der Sluis, Justin, Mirjam Van Praag, and Wim Vijverberg (2008), "Education and Entrepreneurship Selection and Performance: A Review of the Empirical Literature," *Journal of Economic Surveys,* 22(5): 795–841.

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

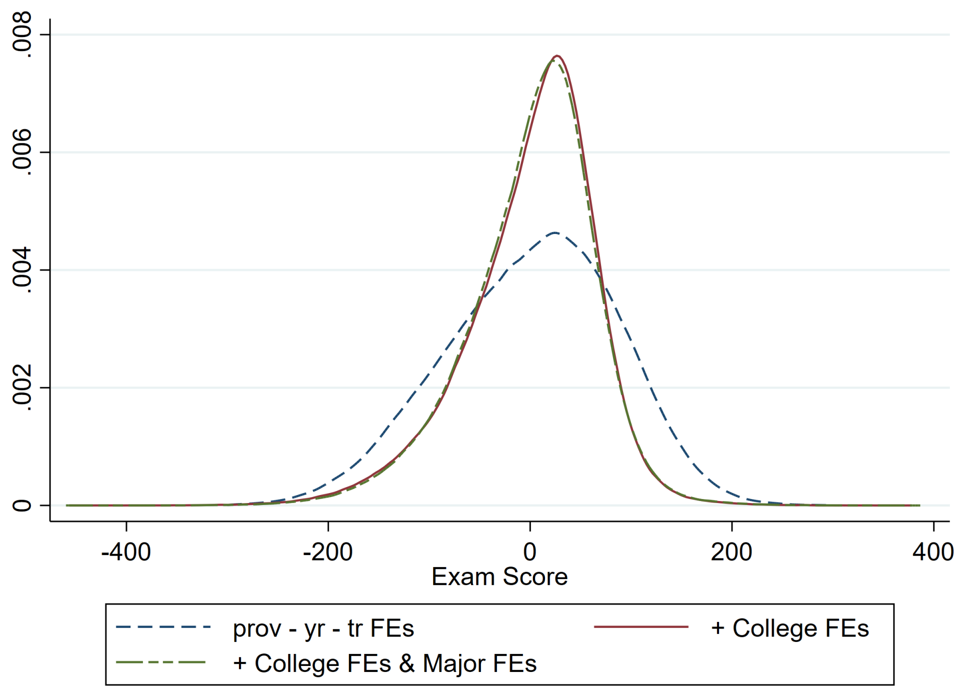
[34] Young, Alwyn (2000), "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China," *Quarterly Journal of Economics,* 115(4): 1091–1135.

Figure 1: Exam Scores Distribution across and within Colleges

(a) Exam Scores by College Ranks



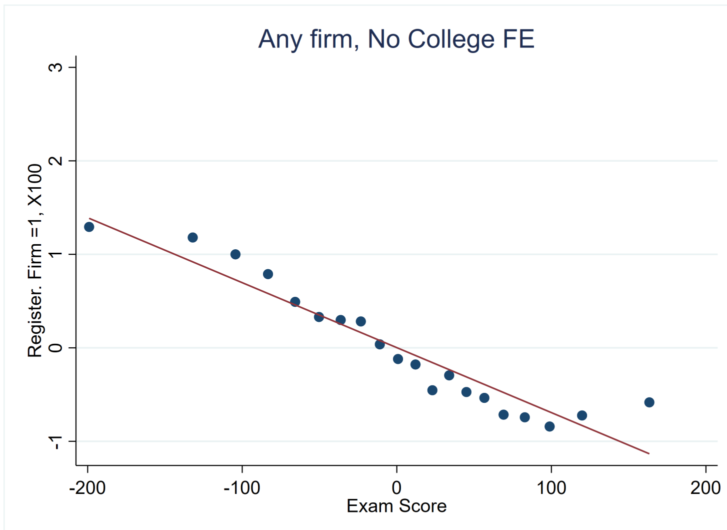
(b) Exam Score Distributions



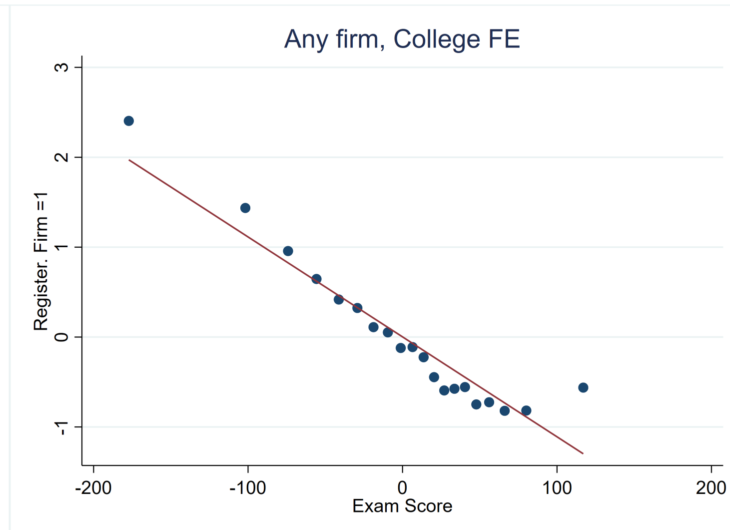
Notes: Figure (a) shows that better colleges have higher-score students on average. Meanwhile, score distributions overlap across different batches of colleges. The two marks outside the box indicates the upper and lower adjacent values. The shaded box ranges from the 25th to the 75th percentiles, where the middle mark indicates the median. Figure (b) confirms the wide variation within colleges.

Figure 2: College Entrance Exam Scores vs. Firm Creation

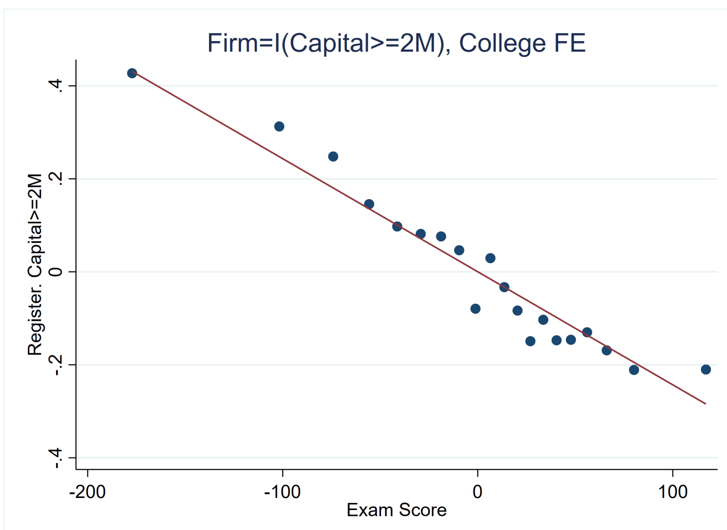
(a) No college FE, any firm



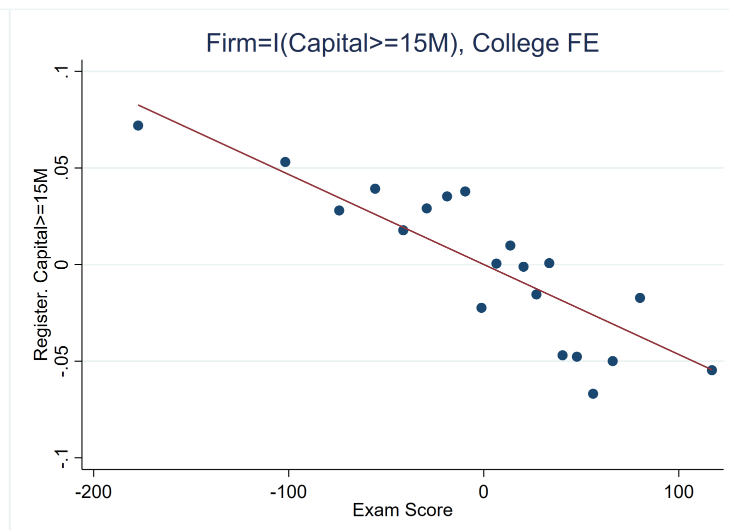
(b) College FE, any firm



(c) College FE, registered capital ≥2M



(d) College FE, registered capital ≥15M

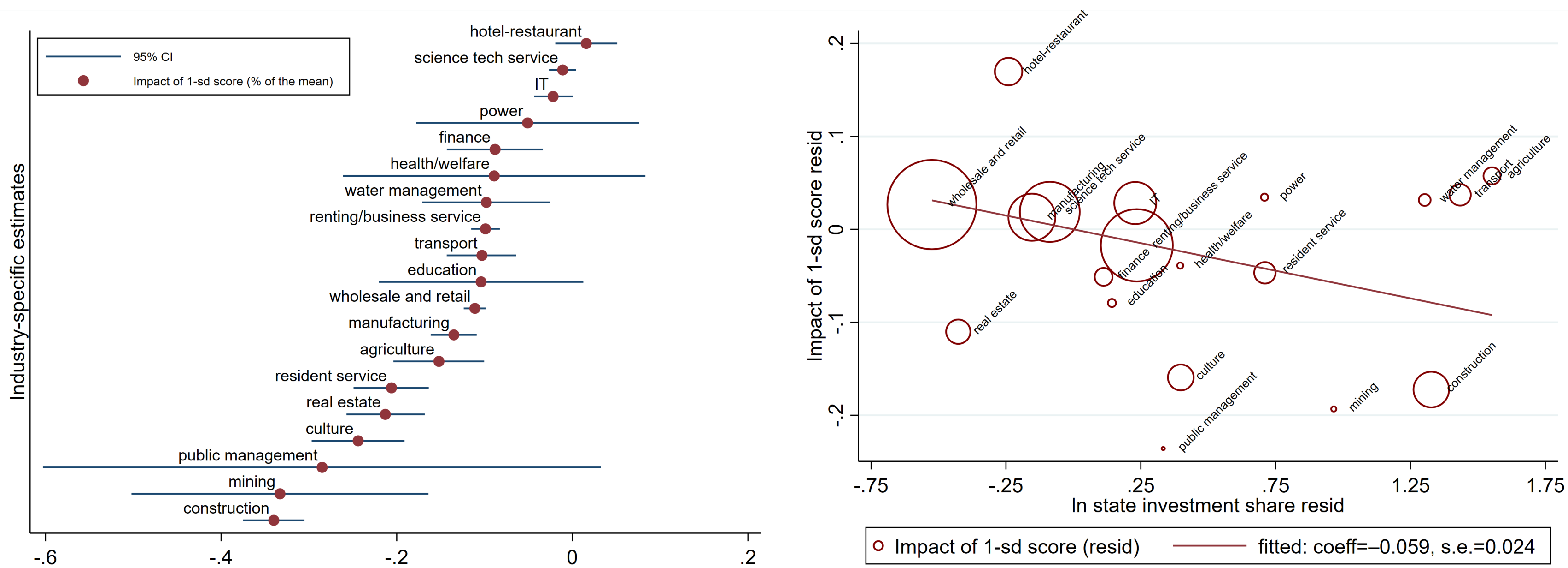


Notes: Firm=I(Capital≥2M/15M) refers to defining a firm only if it is large, i.e., its registered capital no less than RMB 2 million/15 million.

Figure 3: Score-Firm Relationships by Industry

(a) Estimates By Industry

(b) Estimates vs. State penetration (controlling schooling by industry)



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 state 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.

Table 1: Summary Statistics

A. 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
B. 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				
C. 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
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
Birth county ln GDP per capita	8.67	0.81	8.21	9.21	1,814,501

Notes: Our data links administrative data on college admission during 1999–2003 with that on firm registration records from the 1980s to 2015.

Table 2: 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		7.21			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: 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: 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. 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.

Table 4: Scores and Firm Creation by College Rank, 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	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.

Table 5: 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.	$\text{peer}_{c,t}$ (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)
$\text{peer}_{c,t}$ (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 6: Scores and Wage-jobs, within Colleges
Survey Data, Individual-level Analysis

	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
	ln Wage		Jobs providing local Hukou			Relative risk		Relative risk	
	all	all	all	non-local		State sector	Private firms	State sector	Private firms
Exam score (sd)	0.029	0.026	0.037	0.023		1.320	1.189	1.282	1.182
	(0.007)	(0.007)	(0.009)	(0.011)	*p*-value	0.017	0.188	0.042	0.224
Male		0.049	0.108	0.117				1.903	1.166
		(0.007)	(0.009)	(0.010)	*p*-value			0.000	0.246
Urban		0.014	-0.040	-0.005				1.389	1.256
		(0.007)	(0.009)	(0.011)	*p*-value			0.004	0.087
Elite High school		0.019	-0.027	-0.014				1.356	1.472
		(0.008)	(0.010)	(0.012)	*p*-value			0.067	0.047
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
Obs. (excl singletons)	14,801	14,094	14,094	9,731		14,651	14,651	14,094	14,094

Notes: This table presents the correlations between scores and first-job wages and benefits, as well as the relative risk of entering different sectors. The reference group for Columns (4)-(8) is being an entrepreneur.
This data comes from the College Graduate Student Survey we conducted during 2010-2015. Standard errors in the paraphrases are clustered at the college level.

Table 7: Scores and Other Personal Traits, within Colleges
Survey Data, Individual-level Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Academic performance		Political membership	Social activity		Risk aversion (2011)	
	GPA	Academic Awards	Party Membership	Union leader	Union/Social org. leader	Prefer certain pay to a lottery	Ensure certainty in investment
Mean	3.01	0.063	0.265	0.125	0.495	0.636	0.329
Exam score (sd)	0.071	0.016	0.050	0.003	-0.010	0.028	0.039
	(0.012)	(0.007)	(0.019)	(0.007)	(0.014)	(0.031)	(0.028)
Male	-0.278	-0.022	-0.109	-0.019	-0.089	-0.066	-0.036
	(0.016)	(0.006)	(0.024)	(0.009)	(0.013)	(0.029)	(0.023)
Urban	-0.057	-0.001	-0.009	0.035	0.059	0.031	0.002
	(0.013)	(0.006)	(0.017)	(0.009)	(0.012)	(0.034)	(0.039)
Elite high school	-0.002	0.006	0.028	0.032	0.059	-0.015	-0.032
	(0.011)	(0.007)	(0.022)	(0.011)	(0.021)	(0.020)	(0.029)
prov-track-year FE	Y	Y	Y	Y	Y	Y	Y
college FE	Y	Y	Y	Y	Y	Y	Y
age FE	Y	Y	Y	Y	Y	Y	Y
Obs. (excl singletons)	12,226	14,094	14,094	14,094	14,094	3,022	3,022

Notes: This table presents the correlations between scores and various personal traits.
This data comes from the College Graduate Student Survey we conducted during 2010-2015. We only asked risk-related questions in 2011. Standard errors in the paraphrases are clustered at the college level.

Online Appendix

Table of Contents

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

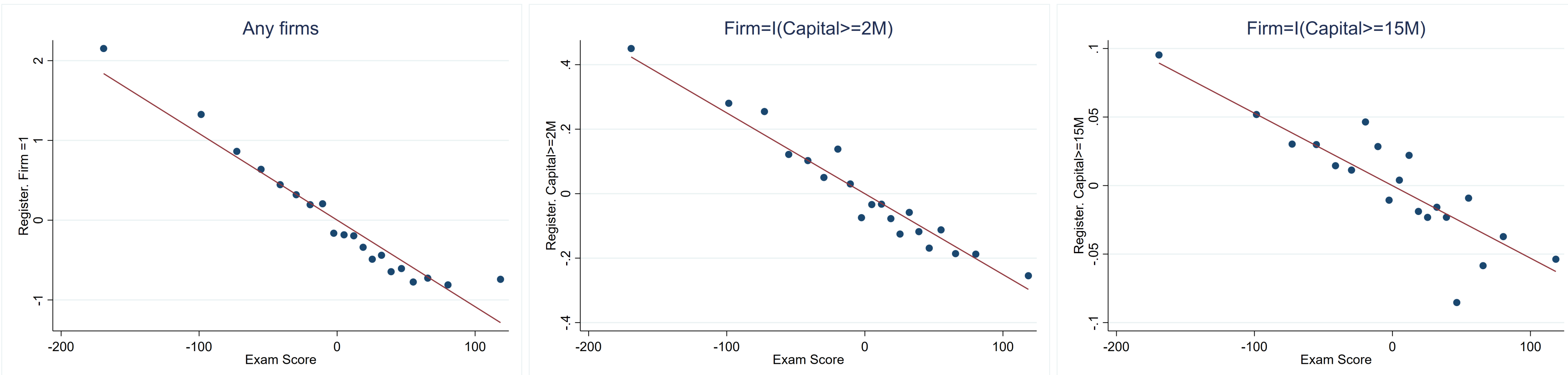
Figure B.1 shows that the negative link between exam scores and firm creation within colleges hold after controlling for 12-major fixed effects. Because major choices reflect one's own preferences, we focus on with-college comparison and report results without and with major fixed effects.

Figure B.1: 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 3.

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 Repeat Exam Takers

Our findings remain similar after excluding repeat 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 Addressing Family Firms

To address the concern of family firms, we keep a subgroup of firms with two conditions: (1) it was established after one took the exam, (2) the age difference between an individual in our college population and the eldest shareholder of the firm is smaller than 20 years. Such a definition is unlikely to include any family firms. The results are presented in Table B.5.

Table B.5: Restricting the Definition of Firms
Administrative Data, Individual-level Analysis

	(1)	(2)
	Firm est. post exam year and age diff among shareholders <20	
mean	5.89	5.89
Exam score (sd)	-0.598	-0.599
	(0.028)	(0.028)
Male		2.984
		(0.056)
Urban		-0.241
		(0.046)
High school quality (sd)		0.064
		(0.026)
ln GDP per capita		0.015
(birth county, 2001)		(0.026)
prov-track-year FE	Y	Y
college FE	Y	Y
age FE		Y
Obs. (excl singletons)	1,814,488	1,814,482

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

C Additional Evidence

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

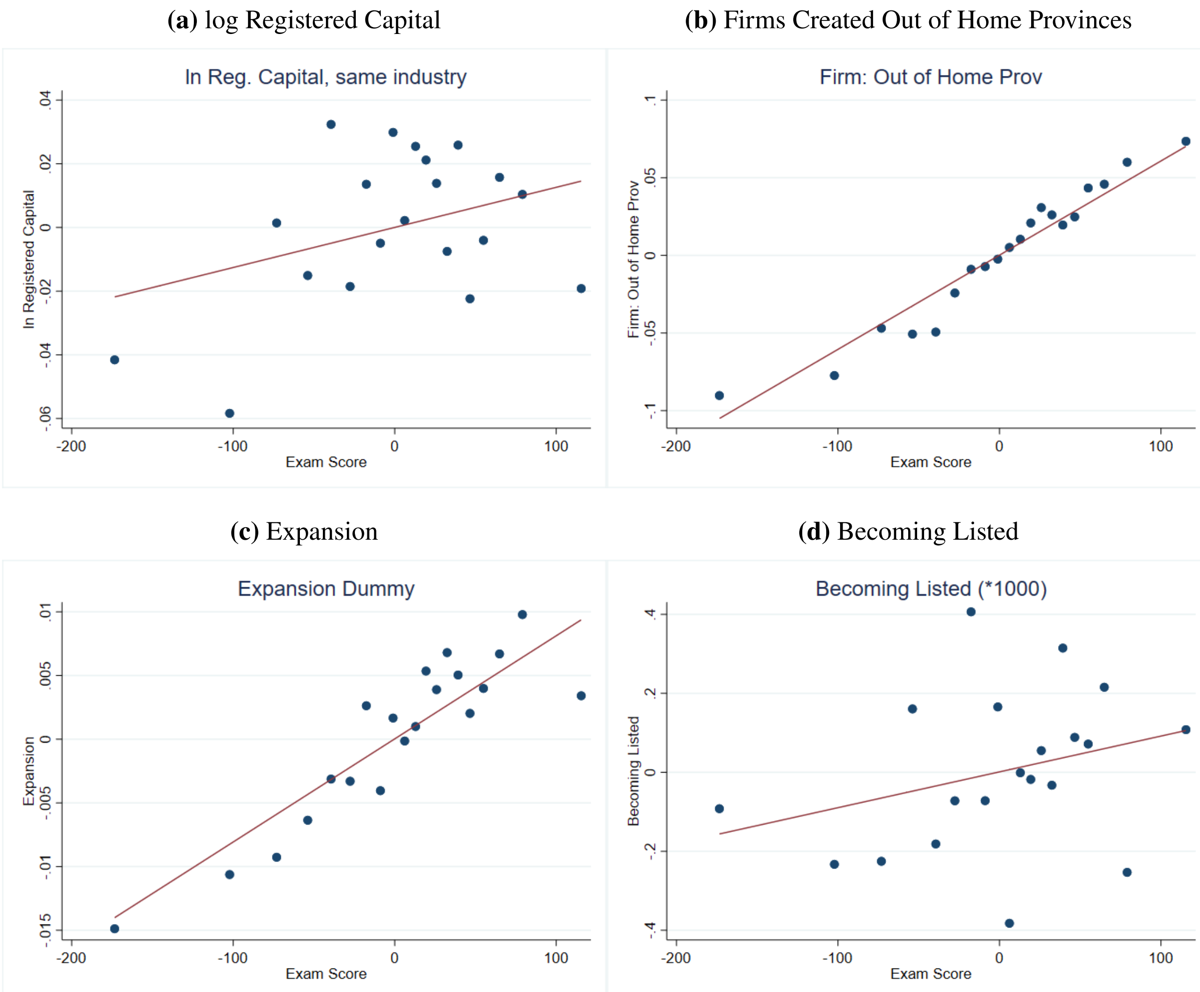
	(1)	(2)
	peer$_{c,t}$ (sd)	
Exam score (sd)	0.0267	0.0006
	(0.0048)	(0.0004)
Male	0.0042	-0.0004
	(0.0038)	(0.0003)
Urban	0.0595	-0.0002
	(0.0045)	(0.0003)
High school quality (sd)	0.0116	-0.0013
	(0.0052)	(0.0003)
ln GDP per capita	0.0066	-0.0005
(birth county, 2001)	(0.0014)	(0.0002)
prov-track-year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

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)



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: 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