Host Favoritism and Talent Selection: Evidence from Chinese Science Olympiads*

Justin Jihao Hong † Xuan Li ‡

This version: February 2024

Abstract

We study favoritism in the selection of elite scientific talent, by examining the relationship between host institution affiliation and performance in the Chinese Science Olympiad, where a gold medal guarantees a student's admission to top universities. Using hand-collected participant-level data (2003 - 2021), we find that students affiliated with the host province have a significantly higher winning probability, and the effect is more pronounced in host provinces where corruption norms are more prevalent. We further present evidence suggestive of cheating behavior using a portion of the contest vulnerable to information leakage, as well as the centralized post-Olympiad selection outside the control of host provinces. Together, our findings shed light on the crucial role of the organizational structure in designing equitable assessment systems for talent.

Keywords: Favoritism, Host Institution, College Admission, Science Olympiad, High-Stakes Test, Talented Youth

JEL: I24, D70, P00

*We are indebted to Raymond Fisman for his continuous guidance and support throughout this project. We would also like to thank Andrew Bacher-Hicks, Chao Fu, Siddharth George, Joshua Goodman, Kevin Lang, Dilip Mookherjee, Daniele Paserman, Johannes Schmieder, and participants at the Applied Micro Workshop (BU 2022) for helpful comments. Yuchen Tao has provided excellent research assistance. We particularly thank Yuzhao Yang and Peiran Xiao for sharing their institutional knowledge. An earlier version of this paper has been circulated under the title "Corrupting the golden path to top universities: rent-seeking in identifying talented youth". All errors are our own.

†Department of Economics, Boston University (email: hjihao@bu.edu)

‡Department of Economics, Boston University (email: xuanli@bu.edu)

1 Introduction

Knowledge creation is critical to long-term economic growth (Romer, 1986). The selection and cultivation of STEM talent play a key role in pushing the frontier of knowledge (Agarwal and Gaule, 2020). To better nurture gifted students, they often receive privileged opportunities, such as greater access to prestigious colleges. However, talent recognition – and thus allocation of opportunities – may be subject to distortions as a result of favoritism and bias (Fisman et al., 2018), which may potentially loom large because of the high stakes and often-opaque selection process.

In admission to top universities, including those in China, talent selection programs such as the Science Olympiad play an important role. The substantial rewards associated with these competitions provide high-powered incentives for students, parents, and educators to compete and succeed. A gold medal ensures that a student will be admitted to a top university.[1] Teachers who place students in top universities may receive monetary rewards and honors. Perhaps driven by these incentives, there have been well-publicized instances of teachers and parents attempting to secure college admission for their students or children through cheating in talent identification programs. For instance, during the 2006 Chinese Chemistry Olympiad, a student was surprised to find that the test questions were almost identical to the practice questions provided by his instructor just one day earlier. In a subsequent poll of 44 instructors, 56 percent agreed that leaks of test questions were very prevalent and have long been an open secret among students and instructors.[2]

Motivated by insights gained from such well-publicized cases, our paper examines the prevalence of biases in high-stakes talent identification programs in China, with an emphasis on the host institution's role. The host institution is an integral element of academic tests and competitions. Just as each International Olympiad has a host country, each national Olympiad in China is organized by a province. Host provinces are responsible for organizing the event, which includes the design of the test, the preparation of lab materials, and the oversight of the contest itself. This setup raises the concern that relationships between the host and its affiliated students might be prone to favoritism. Anecdotes indicate that the host province may exploit

[1] For instance, prestigious institutions like MIT recommend that prospective students compete in Science Olympiads; in 2019, MIT recruited 3 out of the 6 US International Mathematical Olympiad (IMO) team members (source: https://mitadmissions.org/apply/prepare/enrichment/).

[2] See https://news.sina.com.cn/o/2006-09-21/143710076118s.shtml. This type of incident is not an isolated case and can be observed in other non-academic talent identification programs as well. For example, in the notorious Varsity Blues scandal, Devin Sloane falsified his son's athletic profile and paid William Singer $250,000 to secure his son's admission to the University of Southern California as a water polo recruit. While not our focus here, this type of manipulation may also distort talent allocation, as it prevents a more talented student from receiving admission.

its position to gain substantial advantages. For example, during the 28^{th} Chinese Chemistry Olympiad in 2014, there were allegations that Province J, the host, changed the scores of several participants after the contest, enabling two more of its students to gain direct admission to top universities.[3] Another notable example is the 37^{th} Chinese Physics Olympiad in 2020, where 35 students out of 41 from the host province won gold medals, as compared to 120 of 365 participants overall.

Using hand-collected data on participants and host provinces of five core Chinese Science Olympiads spanning over 20 years (2003 - 2021), we show that the host province earns 103.4% more gold medals than when it is not the host. This implies that an additional 2.5 students from the host province are eligible for direct admission to top universities. The host province advantage is highly robust. It survives the inclusion of contest fixed effects, province-subject fixed effects, and province-subject-specific time trends. Furthermore, we implement an event-study specification and find no evidence of pre-trends or lagged effects. These findings suggest that the host province advantage is uniquely tied to the year in which a province assumes the role of host. More importantly, students from the host province demonstrated a significant increase in their chances of winning a gold medal by 20.8% (equivalent to 6.38 percentage points), despite the fact that host provinces have the capacity to send 55.1% additional students (perhaps of lower quality) to the contest as an additional perk of serving as host.

While it is difficult to empirically identify the exact source of host province advantages, we present a number of patterns that suggest possible cheating by the host. Students participating in the Science Olympiad are the best representation of their home provinces and are able to solve challenging questions in a short period of time. There is, however, one important exception: the laboratory experiment. Although maintaining confidentiality of theoretical questions is relatively straightforward, the preparation of experimental materials requires specific equipment and coordination with host institutions, making the process more susceptible to leaks of information. In line with this concern, we find that the experimental part shows a significant performance gap between students from the host province and those from other provinces, especially for gold medal winners. The theoretical part, however, does not exhibit any performance gap. Furthermore, we find that the advantage enjoyed by the host province in the Chinese Chemistry Olympiad (CChO) has almost disappeared since 2020. It was that year that the experimental portion of the competition was removed from CChO as part of a contest reform. We speculate that the removal of the experimental task may have limited the potential for information leakage, and thereby made it more difficult for students from the host

[3] Source: http://zqb.cyol.com/html/2015-01/16/nw.D110000zgqnb_20150116_5-01.htm.

province to leapfrog students from other provinces. While these results can only be taken as suggestive, they are in line with earlier anecdotal accounts that local educators and organizers utilize their informational advantages to support their students. This, in turn, can benefit the careers of related teachers and officials (as there will be more gold medalists and more students admitted to top universities).

In the absence of direct observation of bribery, it is seldom empirically feasible to determine whether the host favoritism is the result of rent-seeking or the host's in-group bias towards its affiliated students.[4] In either case, students from host provinces end up with better performance. In addition to a few well-known cases, we provide heterogeneity analyses that further suggest that rent-seeking may play a role in explaining our results, based on two province-level corruption norm measures: (1) the general perception that "corruption is necessary to achieve success" based on a national representative household survey; (2) a more teacher-specific measure based on the online search index for "gift money" around Teachers' Day, which provides parents with appropriate reasons for giving red envelopes to teachers in China. We find that the host province advantage is more pronounced in provinces where corruption norms are more prevalent, and the estimates remain unchanged when accounting for heterogeneity by GDP per capita and local education investment. Although the presence of in-group biases cannot be ruled out, this finding provides further suggestive evidence in favor of rent-seeking explanations.

Beyond the "host favoritism" explanation, there are two natural alternative interpretations for the higher probability of host-province students winning a gold medal: (1) students feeling more comfortable competing at home and (2) the superior quality of the participants from the host province.

To evaluate the "comfort" hypothesis, we leverage various sources of variation – including host-hometown distances, neighboring provinces, and neighboring city comparisons – to assess performance gaps. The results consistently indicate that the comfort factor is unlikely to play a significant role. To investigate the "quality" explanation, we control for province-year fixed effects to ensure that cohort-specific quality shocks can be accounted for, and we discuss additional channels via province-subject-year changes (strategic reallocation, infrastructure investment, peer effects, etc.). The results also suggest that these factors are unlikely to be significant contributors. Lastly, and more straightforwardly, we show that students from the host

[4] In our context, "rent-seeking" involves players – students, teachers, and organizers – benefiting themselves through favor-exchange and collusion at the expense of other participants. "In-group biases" refer to local organizers and teachers favoring their students, whom they perceive as belonging to the same social group and "insiders," without rent-seeking incentives and bribery.

province are less likely to be selected as National Team Members afterward, resulting from a centralized selection process that cannot be interfered with by local educational authorities (i.e., host provinces). Taken together, we do not find support for either of these alternatives.

Our work contributes first and foremost to the literature on corruption in education (Reinikka and Svensson, 2004, 2005; Ferraz, Finan and Moreira, 2012), especially score manipulation in high-stakes tests (Jacob and Levitt, 2003; Lavy, 2009; Borcan, Lindahl and Mitrut, 2014; Dee et al., 2019). Our study differs from the existing literature in two key ways. First, while college admission fairness serves a critical role in fostering intergenerational mobility (Chetty et al., 2017; Chetty, Deming and Friedman, 2023), corruption in the admission process only surfaces occasionally through qualitative scandals (e.g., the Varsity Blues scandal in which wealthy applicants to U.S. universities bribed school officials or otherwise manipulated the admissions process). We are the first to empirically demonstrate that rent-seeking is prevalent in a nationwide selection process at an elite level, over a long period of time. Second, while existing research often focuses on high school matriculation tests, we delve into a widely used but often overlooked talent selection system: the Science Olympiad. Our work, therefore, differs from documenting score manipulation among students at the lower end of the ability distribution or the entire population (Dee et al., 2019; Jacob and Levitt, 2003; Borcan, Lindahl and Mitrut, 2014).

In a broader sense, we contribute to the ongoing policy dialogue concerning the development of a fair and equitable testing system. While the content of college entrance tests, like the SAT, is generally standardized, scholars have recently discovered that organizational features (like cost of taking the SAT) can exacerbate academic disparities between privileged and underprivileged groups, leading to long-lasting consequences (Bulman, 2015; Goodman, Gurantz and Smith, 2020).[5] Interestingly, we observe that most host provinces are located in more developed areas – a feature that may further exacerbate pre-existing disparities in access to (elite) higher education, and which may be relevant for other contexts (e.g., the location of SAT and language testing centers).

Finally, our paper links to an emerging body of literature that focuses on talented youth (Ellison and Swanson, 2016; Buser and Yuan, 2019; Moreira, 2019; Agarwal and Gaule, 2020; Ellison and Swanson, 2021; Choudhury, Ganguli and Gaulé, 2023). Individuals who excelled in teenage years play a pivotal role in propelling the boundaries of knowledge forward in

[5] Bulman (2015) finds that the opening of a new SAT testing center at a student's own/neighboring high school results in an increase in new test-takers and an increased likelihood of college enrollment. Goodman, Gurantz and Smith (2020) demonstrate that high-income students exhibit a greater tendency to retake the SAT, resulting in improved scores and a higher rate of college enrollment.

the future (Buser and Yuan, 2019). Talent recognition has a significantly positive impact, not only on an individual's own academic performance but also on that of their peers (Moreira, 2019). Our findings shed light on an often-overlooked caveat in the process of identifying talented youth: the screening process relies on high-powered incentives and signals that can be susceptible to favor exchange and biases.

2 Institutional Background

2.1 Chinese Science Olympiads

2.1.1 The Overview of Chinese Science Olympiads

The Chinese Science Olympiads are a group of nationwide annual competitions in various subjects with extremely deep and challenging test questions aiming to select the most talented students to attend the International Science Olympiads. Among all Chinese Science Olympiads, five of them are core subjects: Chinese Mathematical Olympiad (CMO), Chinese Physics Olympiad (CPhO), Chinese Chemistry Olympiad (CChO), China High School Biology Olympiad (CHSBO), and National Olympiad in Informatics (NOI).

Each participant province selects its provincial team members to compete in National Science Olympiads based on students' performance in the Provincial Science Olympiads. There were generally less than 200 participants in each Olympiad in 2003, but the total number of participants substantially increased during the sample period (see Panel A of Figure 1). In 2021, more than 500 students attended CMO. With few exceptions, all participants in the National Science Olympiads are medalists. [6] But the portion of different medalists varies across subjects and changes over time (see Panel B of Figure 1). For example, in the 15^{th} CHSBO (2006), 20% of participants are gold medalists, 30% silver medalists, and 50% bronze medalists; while the portion of gold medalists is 30% in the 25^{th} CHSBO (2016).

The Chinese Science Olympiads for different subjects are organized by different central committees.[7] Moreover, each subject has its own host province each year. For example, in 2003, the host provinces for CMO, CPhO, CChO, CHSBO, and NOI were Hunan, Shandong, Hubei, Sichuan, and Shanghai, respectively. The host province bears the responsibility of organizing the event, encompassing various tasks such as providing food and accommodation,

[6] For example, below the 50% of average score in the 25^{th} CHSBO (2016), below the 25% of total score in the 19^{th} CChO (2006) or cheating.

[7] CMO is organized by Chinese Mathematical Society, CPhO Chinese Physical Society, CChO Chinese Chemical Society, NOI China Computer Federation, and CHSBO is jointly organized by China Zoological Society and Botanical Society of China.

orchestrating the opening and closing ceremonies, and overseeing the competition. Notably, one specific task that recurrently arises is the design of the test. For CMO and CChO, the host province should hire experts or organize a team to design the test. For CPhO, part of the test will be designed by experts recommended by the host province. Furthermore, in the case of Olympiads involving laboratory experiments (such as CPhO, CChO, and CHSBO), the host province is required to procure and prepare specialized laboratory equipment and instruments tailored to the experiments to be conducted during the Olympiads. As a reward for the contributions made by the host province, the central committee allocates additional quotas to them. Consequently, the host province can typically send a greater number of students to participate in the Olympiad. For CHSBO, the host school can even send its own team to compete in the Olympiad.

2.1.2 The Determination of the Host Province

All provinces interested in acting as the host province for a particular Olympiad must submit the application to the central committee no later than one year (usually two years) prior to the event. The central committee will make the final decision and select one host province from all candidate provinces mainly based on organizational ability and fiscal support. For some subjects that involve laboratory experiments (i.e., Physics, Chemistry and Biology), high-quality lab equipment and lab instruments are a must. Moreover, all other things being equal, the central committee will give priority to these provinces that have not been a host province or have won gold medals in the International Science Olympiad in recent years.[8] As a result, the host province status is neither randomly assigned to participant provinces nor is rotated among participant provinces. Some provinces may be particularly strong at some subjects or may attach more importance to Science Olympiads and thus are more motivated or qualified to organize Science Olympiads. As Figure 2 shows, some provinces like Zhejiang, Hunan, and Sichuan frequently serve to be host provinces, while some provinces have never acted as host provinces in the last twenty years.

2.2 High-Powered Incentives

Although the main purpose is to select the most talented students to represent China in attending the International Science Olympiads, Chinese Science Olympiads are closely related to college admission. Top universities (such as Tsinghua University and Peking University) will dispatch an admissions team to directly recruit top-performing students from Chinese Science

[8] The information on candidate provinces and the exact formula for deciding the host province are unavailable.

Olympiads. Thus, students and local education agents are strongly incentivized to stand out in Chinese Science Olympiads.

High Schoolers. *Gaokao* (National College Entrance Exam) is the foundation of the college admission in China, and only less than 0.1% of students gained college admission without sitting for *Gaokao* in 2010 (Bai, Chi and Qian, 2014). However, admission to top-notch universities is a completely different story. Among all students admitted to top two universities in 2021, only less than half of them were admitted based on the raw score in the *Gaokao*.[9] Gold medalists in the Olympiads will have the opportunity to be admitted through the recommendation track, which means they can gain admission to prestigious universities without the need to take regular entrance exams.[10].

To illustrate the importance of the Chinese Science Olympiads in admission to the top two universities, we take Sichuan Province as our example. In 2013, less than 90 students in Sichuan Province could enter Tsinghua University or Peking University based on the raw scores in the *Gaokao*. However, these two universities admitted more than 290 students from Sichuan. Specifically, in 2013, Peking University's admission cutoff in Sichuan was 690 points. Less than 50 students earned more than 682 points. However, if a student earned 60 bonus points through the Chinese Science Olympiads, she only needed to be ranked 3,243 in Sichuan to enter Peking University.

Local Education Agencies and Teachers. In the *Gaokao*, each university typically allocates a fixed admission quota for students from each province. However, gold medalists in the Chinese Science Olympiads have the opportunity to be admitted to top-notch universities through an alternative track. These students, admitted through the recommendation system, do not count towards the regular *Gaokao* quota. As a result, a province can send more students to prestigious universities like Tsinghua University and Peking University if it performs well in the Chinese Science Olympiads, as the gold medalists bypass the usual admission quotas. Given the pivotal importance of college entrance outcomes in the educational system, the Olympiad performance can significantly impact the careers and reputations of education-sector officials and the teachers of the participants. Prominent teachers of gold medalists often receive invitations to deliver lectures in other high schools and specialized training institutions. Further-

[9] See https://www.sohu.com/a/500146640_120828047.

[10] To curb the rampant rent-seeking behavior in the recommendation process, the central government significantly reduced the size of the recommendation track in 2013 (source: http://edu.people.com.cn/n/2015/0513/c1053-26995459.html). After 2013, only top-ranked gold medalists (i.e., national team candidates) can be admitted through the recommendation track. But others can generally earn bonus points (up to 60 points) in the *Gaokao*, and some can even be admitted to Tsinghua/Peking University as long as they pass the admission line of the first-tier university.

more, private high schools offer highly competitive compensation packages to attract these exceptional instructors. As a case in point, Tianli High School in Chengdu offers salaries reaching up to ¥3,000,000 (approximately $400,000).

3 Data and Identification

3.1 Data

We compile a comprehensive dataset of the Chinese Science Olympiads from 2003 to 2021, tracing 92 contests in which the information on contest locations and participants is available.[11]

Data on the host province are collected from the official website of Chinese Science Olympiads maintained by China Association for Science and Technology.[12] During the sample period, we have 22 unique host provinces and 36 unique host cities. As can be seen in Figure 2, there is substantial variation in the frequency of being the host province. The top three are Zhejiang (13), Hunan (10), and Sichuan (8). Provinces that have never acted as host provinces in the last twenty years are concentrated in Western China. Four provinces – Zhejiang, Hunan, Jiangsu, and Fujian – host all five Chinese Science Olympiads, and six provinces host four of them. Some provinces host one particular Olympiad more than once during the sample period. A case in point is Zhejiang, acting as the host province for NOI in 2008, 2015, 2017, and 2021.

Medalists' information for the Chinese Science Olympiads is mainly obtained from (1) the official website of China Credential Verification maintained by the Ministry of Education, and (2) the Service Platform of Science and Technology Innovation Activities for Children and Youth maintained by China Association for Science and Technology, where these data have been published since 2003.[13] Supplemental data on medalists are collected from other official resources and various newsletters, including the official website of each Olympiads maintained by their central committees, and CPhO's official publication *Annual CPhO Bulletin*.[14] Our final sample contains 21,643 medalists from 976 high schools, including information on name, high school, and award (i.e., gold, silver, or bronze medal).

The medalist lists that we obtained from the above sources are used to construct our four main outcome variables. Three are at the province-contest level (*Proportion of Gold Medals*, *Proportion of Participants*, and *Prob. of being Gold Medalists*), and one is at the individual level

[11] Medalists information for 22^{nd} CMO (2007), 25^{th}, 26^{th} CChO (2012), and 20^{th} CHSBO (2011) are not available.

[12] https://cso.cyscc.org/.

[13] https://www.chsi.com.cn/mdgs, https://gs.cyscc.org , and http://www.xiaoxiaotong.org.

[14] For example, CMO (http://www.cms.org.cn), CHSBO (http://www.botany.org.cn and http://czs.ioz.cas.cn/swxjs/qgjs), and NOI (https://www.noi.cn/hjmd/mdgs).

(*Gold Medalist*). *Proportion of Gold Medals* is the number of gold medalists a province earns in a given contest, divided by the total number of gold medals in that contest.[15] *Proportion of Participants* is the number of participants from a province in a given contest, normalized by the total number of participants in that contest.[16] *Prob. of being Gold Medalists* is the number of gold medals a province earns, divided by the number of participants from that province.[17] Since there are 31 participant provinces in most cases, the mean of *Proportion of Gold Medals*, and *Proportion of Participants* are both around 0.033 by construction.[18] *Gold Medalist* is a dummy variable indicating whether a participant wins a gold medal. On average, 26.6% of participants would win a gold medal. To have a better sense of the magnitude of the host province advantage in our context, we generate two additional outcome variables: *# Gold Medals* is the number of gold medalists a province earns in a given contest, and *Number of Participants* is the number of participants from a province in a given contest.

By combining city of high school information on the medalist and the host province, we generate the province-contest (or sometimes participant) level variable $HomeProv_{pc}$, denoting that province p is the host province of contest c. Similarly, we generate three participant-level variables $HomeCity$, $SameProvOtherCity$, and $NeighboringProv$, denoting whether a student is from the host city, from the host province but not the host city, or from neighboring provinces of the host province, respectively. These three variables capture the extent to which feeling more comfortable can explain the gold medal premium for students from the host province. Moreover, to investigate the impact of travel and acclimatization on performance, we generate $Distance$ by calculating the distance between one's home city and the host city, measured in 100 kilometers. To test the potential spillover effect among subjects, we additional generate *HomeProv_of_OtherSubj*, a dummy variable that is one if the participant is from a province that hosts any academic Olympiad of other subjects in that year.

To better understand how gold medal premium occurs, we collected detailed score data for theory part and experiment part in three CHSBOs.[19] The dependent variable is *Standardized Score*: the test score standardized within the contest. We generate a dummy variable *Lab Experiment*, which equals one if the outcome corresponds to the test score of the lab experiment part, and 0

[15] $\text{Proportion of Gold Medals}_{pc} = \frac{\text{\# Gold Medalists from Province P in Contest C}}{\text{\# Gold Medalists in Contest C}}$.

[16] $\text{Proportion of Participants}_{pc} = \frac{\text{\# Participants from Province P in Contest C}}{\text{\# Participants in Contest C}}$.

[17] $\text{Prob. of being Gold Medalists}_{pc} = \frac{\text{\# Gold Medalists from Province P in Contest C}}{\text{\# Participants from Province P in Contest C}}$.

[18] Tibet has never participated NOI, CHSBO, and CChO. Additionally, Qinghai has never attended NOI.

[19] Test scores for 8^{th}, 16^{th}, and 21^{st} CHSBO are obtained from https://cso.cyscc.org/biology, https://cso.cyscc.org/AttachFile, and http://www.botany.org.cn, respectively.

if the outcome corresponds to the test score of the theory part.

Moreover, to understand the relationship between prior measured ability and performance in the National Science Olympiads, we collect additional data on gold medalists in Provincial Science Olympiads from the official website of China Credential Verification maintained by the Ministry of Education.[20] From 2003 to 2006, the medalists are ranked by their performance in the name list. Thus, we are able to construct provincial team members' rankings in both Provincial Science Olympiad (*Ranking in Provincial Pre-Selection*) and National Science Olympiad (*Ranking in National Olympiad*).

Lastly, to assess the quality of participants from the host province, we compiled a supplementary dataset on the name list of national team candidates and national team members from 2013 to 2021.[21] We generate two participant-level outcome variables. *National Team Candidate* indicates whether a gold medalist is ranked high enough in the National Science Olympiads to be included as a national team candidate. *National Team Member* is a dummy variable that equals one if a gold medalist is selected as the national team member to represent China to compete in International Science Olympiads. National team members are extremely selective: only 1.4% of participants in the National Science Olympiads would be selected as national team members.

Table 1 provides summary statistics for the full sample. Note that $HomeProv$ is a province-contest level variable when the analysis is conducted at the province-contest level and is an individual level variable when the analysis is conducted at the individual level. The mean of province-contest level $HomeProv$ is 0.33 by construction since only 1 out of 31 participant provinces would act as the host province for a particular Olympiad. The mean of individual level $HomeProv$ is higher (i.e., 0.065) due to the quota bonus rewarded to the host province.

3.2 Empirical Strategy

Our analysis relies on both province-contest-level and individual-level identification. At the province-contest level, we estimate the following specification:

$$Y_{pc_{(st)}} = \alpha_c + \beta * HostProv_{pc} + \pi_{ps} + (\pi_{ps} * f(t))'\theta + \epsilon_{pc} \quad (1)$$

[20]https://www.chsi.com.cn/mdgs.

[21]Data on the national team candidate are obtained from the Service Platform of Science and Technology Innovation Activities for Children and Youth (https://gs.cyscc.org). Data on the national team member can be easily obtained online.

where $Y_{pc_{(st)}}$ is the outcome of interest; p indexes the province, c indexes the contest (subject-year), s refers to the subject, and t is the time (year).

$HostProv_{pc}$ is a dummy variable that is 1 if province p is the host province for contest c in year t. β is the coefficient of interest that captures the average impact of acting as the host province. α_c denotes the contest (subject-time) fixed effects, which control for all contest-specific differences across contests (e.g., subjects, the total number of participants, the form of contest, etc.), enabling a clear interpretation of our estimate (within-contest comparison). Importantly, we further introduce π_{ps}, the province-subject fixed effects, to account for the possibility that a province may be particularly strong in some subjects and thus is more likely to apply for being the host. Given the time span of our sample (19 years), we also include the province-subject-specific linear time trends in our baseline specification: $\pi_{ps} * f(t)$. The inclusion of the province-subject-specific time trends can deal with the potential endogenous selection of the host province due to the priority given to the province with an excellent performance record in recent years. Finally, standard errors are clustered at the province-subject level, since we exploit variation at the province-subject-year level. We also demonstrate robustness by using other clusters in Appendix.

To understand the impact conditioning on participation, we also replicate the identification using the participant-level data:

$$Y_i = \alpha_c + \beta * HostProv_{pc} + \pi_{ps} + (\pi_{ps} * f(t))'\theta + \epsilon_{pc} \quad (2)$$

, where i denotes the individual participant from province $p = p_{(i)}$. All the remaining variables are defined similarly as before.

4 Main Results

4.1 Baseline: Host-Province Advantages

We begin by establishing the relationship between acting as the host province and performance in the contest, using the province-contest level specification. The outcome variable is the "proportion of gold medals" – the number of gold medalists a province earns in a given contest, normalized by the total number of gold medals in that contest. The greater the share of gold medalists a province possesses, the more its high schoolers gain access to top-notched universities via the recommendation track (relative to other provinces).

Table 2 presents the results. There are significant performance differences between provinces that have ever hosted (hereinafter referred to as "ever-host provinces") and those that have never hosted (hereinafter referred to as "never-host provinces"). In fact, ever-host provinces on average earns 4.42% of gold medals, and never-host provinces only account for 0.48% of gold medals, speaking to the potential endogenous application for being the host province as suggested by Figure 2. Therefore, we include both contest fixed effects and province-subject fixed effects in Column (1) of **Panel A**. The inclusion of province-subject fixed effects helps us to absorb fundamental differences in subject-specific performance across provinces. We document a significant host province advantage in Chinese Science Olympiads.

Column (2) of **Panel A** presents the result using our preferred specification, further controlling for province-subject-specific linear time trends. The estimates remain virtually unchanged: on average, a province earns 103% more gold medals when it acts as the host province compared to the period when it does not host the event. To provide a better sense of the magnitude in our context, Columns (3) and (4) replicate the exercise using the number of gold medals a province earns in a given contest. Acting as the host province, on average, is associated with earning 2.5 additional gold medals (and thus 2.5 more high schoolers on a fast lane to top-notched universities).

We also investigate the pattern by subject in **Panel B** of Table 2. All subjects feature significant host province advantages, except for Mathematics (CMO). As shown later, the host province of CMO receives the least generous quota bonus; meanwhile, Math is widely regarded as the most rigorous one among the five subjects.[22]

4.2 Participation Rates

Given that most contests reward the host province by assigning certain quota bonuses (i.e., the host province may send more participants), a natural starting point is to perform a reality check to examine the role of this mechanical channel.

Table A1 presents the estimation results. On average, host provinces can send about 4.4 additional participants (or 55% more participants). Columns (3)-(7) further examine the quota bonus by subject. There are two patterns. First, the participation rate varies by subjects.[23]

[22]First, unlike other subjects with a lab experiment part, CMO only has an intense written test (with four extremely challenging problems). Second, the nature of Math ensures relatively objective grading since all answers and solutions can be easily verified.

[23]Each subject may have its own distinct approach for determining the number of participants allowed from different provinces. For instance, in the case of CHSBO, an evenly divided approach is adopted. This means that the total number of participants is divided equally among the provinces. In contrast, for other subjects, the quota assignment process may take into account additional factors such as the population of the province, its participation rate in provincial Olympiad, and its past performance in the national and international Olympiad.

When not hosting the CHSBO, ever-host provinces exhibit a similar participation rate to that of never-host provinces (i.e., 0.0341 v.s. 0.0308). While the difference in participation rate between ever-host provinces and never-host provinces is quite large in the NOI (i.e., 0.0532 v.s. 0.0209). Second, although host provinces of all five subjects exhibit an increased participation rate, the quota bonus varies by subjects substantially. The host province of CHSBO can send almost an extra team to participate (97.9% more participants). Additionally, the quota bonuses allocated to the host province of CMO, CPhO, CChO, and NOI are on average 13.7%, 63.4%, 27.1%, and 60.0% more participants, respectively. Therefore, the host province advantage we document can be partly attributed to the institutional feature of quota bonuses, as some of the additional participants from the host province have the potential to earn a gold medal.

4.3 Probability of Winning a Gold Medal

As the number of gold medals is determined by the product of the number of participants and the probability of winning a gold medal, we explore another potentially important factor in this subsection: the likelihood of winning a gold medal (conditional on participation).

The likelihood of winning a gold medal might be negatively associated with the number of participants. Pettigrew and Reiche (2016) document that in the Olympic Games, the host country generally can send more athletes due to the less strict qualification requirement for them, but meanwhile, the probability of winning a medal is slightly lower as the average quality of the athletes decreases. Similarly, since the host provinces receive quota-based bonuses resulting in additional (less-qualified) participants, we may anticipate a lower probability of winning the gold medal for students from the host province (i.e., negative estimated coefficients). This effect should be particularly pronounced for subjects with more generous quota bonuses.

Before digging into participant-level data, Column (1) of Table 3 uses province-contest level specification (the same as that in Table 2) to examine the empirical pattern. The outcome variable is the share of gold medalists from a province – the number of gold medals a province earns, normalized by the number of participants from that province. Acting as the host province is associated with a 6.38 percentage-point increase in the average share of gold medalists, suggesting that even if conditioning on the number of participants, the host province still enjoys a discernable gold medal premium. Column (2) of Table 3 further corroborates the analysis by utilizing participant-level data pooling all contests. The magnitude of the estimated coefficient is consistent with that in Column (1).

In Columns (3) - (7), we examine the conditional probability of winning a gold medal by subject. All estimated coefficients are positive, indicating there is non-decreasing probability of earning a gold medal when the individual is from the host province. It is worth noting that the coefficient for Chemistry (CChO) is statistically significant and has a considerable magnitude. Do these results imply the gold medal premium exists only for CChO? One missing part in understanding the observed unvaried probability of winning a gold medal is the increased participation rate. In other words, the observed unvaried probability of winning a gold medal could potentially be an increased probability compared to the counterfactual – as long as the average quality of additional participants is lower in the presence of quota bonuses. In section 5.3.1, we provide evidence to show that the average quality of additional participants is likely to be lower. This suggests that there are potentially two counteracting forces contributing to the probability of winning a gold medeal: (1) the presence of less-qualified additional participants, which decreases the probability of winning a gold medal, and (2) other yet-to-be-identified factors (e.g., cheating), which may increase the likelihood of winning a gold medal. From the perspective of this explanation, the larger and positive magnitude observed in CChO may be attributed to the fact that the influence of the first force is relatively weaker in CChO compared to other subjects – the host province of CChO can send fewer additional participants.

Lastly, Table A2 summarizes the main results thus far with different clustering criteria (e.g., province-subject, province-year, province levels) and small-cluster adjusted inferences. Changing the level of clustering does not affect our statistical inferences.

4.4 An Event-Study Design

In this subsection, we adopt an event-study specification to transparently illustrate our baseline findings thus far, and investigate whether there exists any significant lead-lag effect. Specifically, we estimate the following equation (taking the province-contest level analysis for example):

$$Y_{pc} = \alpha_c + \sum_{j \neq -1} \beta_j * \mathbf{1}(\mathit{Years\ since\ Host}_{pc} = j) + \pi_{ps} + (\pi_{ps} * f(t))'\theta + \epsilon_{pc} \quad (3)$$

, where $\mathbf{1}(\cdot)$ is an indicator function, and $\mathit{Years\ since\ Host}_{pc}$ refers to the event time relative to the year when the province p hosts the contest c. We take $j = -1$ as the omitted baseline group. The remaining variables are defined as before. Similarly, we could replicate the event-study exercise using the participant-level data.

Figure 3 visualizes the results. In each panel, there is a distinct jump in the outcome of interest in the host year relative to other years. This empirical pattern implies that conditioning on covariates from our preferred specification, the timing of acting as the host province is unlikely to be systematically associated with the remaining unobserved province-subject-specific characteristics in the past few years. There is a slight lagged effect in terms of the participation rate, because the quota assigned to each province is (weakly) positively correlated to its past performance.

5 Channels for Increased Winning Probability

Generally, three potential channels are underlying the increased probability of winning a gold medal for students from the host province: host favoritism (with cheating behaviors), comfort, and better quality. In this section, we provide a collection of suspicious patterns that align with the cheating explanation, and further show that feeling more comfortable and better quality are unlikely to be the dominant factors at play.

5.1 Suggestive Evidence on Cheating

5.1.1 Substantial Experiment Score Premium

A laboratory experiment is an inseparable part of science. The practical part consisting of experimental problems forms an important part of the CPhO, CChO and CHSBO, accounting for 50% of the total points. The experimental test requires students to conduct complicated laboratory experiments in a short time (4-6 hours), which generally involve more than 20 different types of equipment, various and potentially hazardous chemicals, and tricky technical procedures.[24]. Thus, a new experiment is prone to failure by its very nature.[25] Given the intricacies involved, even an experienced experimenter may encounter significant challenges when attempting to carry out a new experiment flawlessly within a limited time frame, particularly if they lack prior experience in conducting similar experiments. Furthermore, students are not only evaluated by the final outcome; they are expected to succeed in the first attempt. Restarting a failed attempt is associated with a 20-point penalty out of 100 total points.

The organization of the laboratory experiment in the competition requires great effort made by the host province, and it is inevitable for the host province to know the specific labo-

[24] To have a better sense of the complexity involved in the the experimental problems, past problems for International Biology Olympiads and Intentional Chemistry Olympiads can be found on their official websites: https://www.ibo-info.org/en/info/papers.html and https://www.icho.sk/competition-problems/. Also, *Analysis of the Experimental Part of the IChO Competition* written by Anton Sirota is a useful resource to understand the experimental problems in IChO.

[25] In preparation for the practical part in the CChO and CHSBO, school instructors will often draw students' attention to important phases and illustrate the tricky components in great detail to increase the success rate. However, in the Olympiads, students have to rely on themselves.

ratory materials and equipment involved in the test in advance. Moreover, the host province is often highly engaged in the design of test questions.

Would host-province students somehow be better positioned to navigate the complexity involved in the new laboratory experiment and leapfrogging their competitors? A comprehensive breakdown of scores for each segment of the contest is typically unavailable, with the exception of a few CHSBOs. Taking advantage of these records, in Table 5, we test whether students from the host province can have better performance in the experiment part of the contest. There is no performance gap in the theoretical part between the host province and other provinces. However, students from the host province substantially outperform other students in the laboratory experiments, which is especially pronounced for gold medalists.[26] These findings suggest that host-province students are somehow well-prepared for the practical part and are able to conduct the complicated experiments smoothly even under great pressure within a short time.

5.1.2 Elimination of the Experiment Tests: Chemistry Contest Reform

What might occur in the absence of an experimental component? The 2019 Chemistry Contest Reform presents us with an intriguing shock. In 2019, the central committee decided to cancel the laboratory experiment in the CChO starting from 2020.[27] In Figure 4, we illustrate the probability of winning a gold medal for the host province of CChO by year. It is worth mentioning that a single positive coefficient does not necessarily suggest anything suspicious for a particular host province: for instance, a host province could coincidentally have a better cohort in the year it organizes the event.[28] But interestingly, the coefficients are almost all positive during the whole sample period but suddenly became negative in 2020 (34^{th} CChO), the year the reform was implemented. Importantly, a negative coefficient is consistent with the argument that sending additional participants could drive down the probability of winning a gold medal due to a lower quality of the additional participants (we will discuss this in great detail later). One may concern that it could also be the case that host provinces in 2020 and 2021 are special cases: they would have a negative coefficient even in the absence of the reform. This concern can be partially alleviated by the large positive coefficients exhibited when they acted as the host province for 16^{th} and 23^{rd} CChO.

Taken together, although we cannot take a strong stance on the specific mechanism in-

[26] If the transparency of test scores is endogenous, our results are likely to provide a lower bound.

[27] The reform also comes with two other major changes: (1) the answer will be graded by two referees independently, and (2) students can check their scores online. Therefore, our results can at best provide suggestive evidence on the impact of the removal of the experiment part.

[28] We later provide quantitative evidence suggesting that host-province medalists are actually less qualified.

volved, the accumulated evidence from the experiment part and the reform provides at least some suspicious patterns suggestive of cheating. In the following sections, we will thoroughly examine alternative explanations: feeling more comfortable and better quality.

5.1.3 Further Results: Heterogeneity by Host Province Corruptness

Given that teachers care about their students' futures (Dee et al., 2019), the cheating associated with host favoritism can be interpreted as in-group biases towards local students . Nevertheless, high-powered incentive schemes in education can also lead to behavioral distortions among teachers, including the manipulation of students' test scores (Jacob and Levitt, 2003). For instance, the Atlanta testing scandal infamously revealed that Dr. Hall received more than $500,000 in performance bonuses for her students' superior performance through the erasure of incorrect answers and the correcting of them.[29]

Similarly, in China, teachers who help students outperform in Science Olympiad and place them at top universities will receive honorary and monetary rewards. Teachers of gold medalists are often invited to present lectures at other high schools and specialized training programs as a result of their achievements. For these exceptional instructors, private high schools offer highly competitive compensation packages. In Chengdu, Tianli High School offers salaries up to ¥3,000,000 (approximately $400,000).

Borcan, Lindahl and Mitrut (2014) show financial incentives can lead teachers to participate in corrupt activities like score manipulation. High performance rewards (for both students and teachers) associated with Science Olympiads naturally raise concerns about rent-seeking behavior. Since rent-seeking is hidden by nature, it is hard to sysmetically document parents directly bribing instructors or instructors exchanging favors with organizers. Nevertheless, there are still few high-profile exceptions. For example, according to official media reports, during the 2005 Inner Mongolia Autonomous Region Olympiad in Informatics, some parents managed to obtain the exam questions for their children by purchasing them from the professor responsible for designing the exam, paying ¥100,000 (approximately $14,000) for the confidential information.

To make some empirical progress, we examine whether the extent of host province advantage varies across provinces based on their prevalence of corruption norms. We develop two province-specific measures. As a first measure, the average perception of corruption in each province is evaluated based on the China Family Panel Studies (2010), a national representative

[29] https://www.nytimes.com/2013/03/30/us/former-school-chief-in-atlanta-indicted-in-cheating-scandal.html.

survey of households. Second, we develop a more teacher-specific measure based on the search index for "gift money" (*Hongbao* in Chinese) around Teachers' Day, normalized by the same index for the entire year. In China, it is customary for some parents to present red envelopes to teachers on Teachers' Day as a token of gratitude (and to some extent favor-exchange), asking for additional care and attention teachers provide to their own children.[30]

Table 4 presents the results. we consistently find that the host province advantage is significantly more pronounced in provinces with higher corruption norms. Moreover, when accounting for GDP per capita and local education investment, the estimates remain virtually unchanged. The present findings provide additional empirical support for the rent-seeking explanation beyond anecdotes, although pure in-group biases cannot be completely excluded.

5.2 Alternative Explanation: Comfort

China's vast expanse encompasses diverse cultural, climatic, and dietary factors. Extensive travel for exams can expose students from non-host provinces to unfamiliar environmental conditions. Hence, students from the host province may excel simply due to enhanced comfort.

Table 6 explores this channel. We first investigate whether the probability of winning a gold medal is associated with the continuous distance between a student's home city and the host city. Column (1) shows that, once holding the host province status constant, the coefficient of distance is negligible and not statistically different from zero. This implies that travel and climate/diet differences barely influence the performance of these talented students in high-stakes exams.

Second, we dis-aggregate participants into four discrete groups: students from the host city, students from the host province but not the host city, students from neighboring provinces of the host province, and others (baseline group). Students from the host province but not the host city still need to take a short-distance trip to the host city. Similarly, students from a neighboring province also only need to travel a short distance, and the difference in climate or diet should be small if any. Thus, if comfort plays a prominent role in high-stakes exams, we would expect that students from the host city outperform the other three groups, and the performance gap between students from a neighboring province and students from the host province but not the host city should be small. Column (2) of Table 6 reveals that students from the non-host cities within the host province perform equally as well as those from the host

[30] In response to the widespread practice of gifting red envelopes to teachers on Teachers' Day, numerous local governments and schools have implemented policies requesting teachers to return these red packets after Teachers' Day.

city. However, these host-province students significantly outperform their counterparts from neighboring provinces.[31]

Lastly, we adopt a more stringent specification by comparing students from contiguous city-pairs that straddle the province border. This design is built on comparisons among local areas that are contiguous and similar, except for the host province status. We consider two cities to be a pair if they share a province border and belong to different provinces.[32] As shown in Column (3) of Table 6, students from the host province still substantially outperform students from the adjacent city but belong to a different province. Taken together, these findings provide evidence against the "comfort" explanation as the primary channel.

5.3 Alternative Explanation: Quality

5.3.1 Quota Bonus and Quality of Participants

The host provinces all have quota bonuses. As noted in the discussion of Table 3, if the average quality of additional participants is lower, a non-decreasing (increasing or unvaried) winning probability of host-province participants can indicate the presence of other favorable factors at play (Pettigrew and Reiche, 2016).[33] We thus start with examining the quality of extra participants sent by host provinces.

Before the National Science Olympiad, each province typically selects participants (referred to as "provincial team members") based on their score rankings in the Provincial Science Olympiad. Due to quota bonuses, host provinces can send lower-ranking members to participate in the National Olympiad. We collect data on each provincial team member's performance in both Provincial Science Olympiads and National Science Olympiads. The comprehensive data on Provincial Olympiad rankings is available from 2003 to 2006.[34] The raw correlation between Provincial Olympiad rankings and (within-province) National Olympiad rankings is

[31]It is also worth noting that students from neighboring provinces exhibit a slight advantage over students from a distant province (referred to as the baseline group). This disparity may likely be attributed to the fact that students from under-performing provinces, such as Tibet, are frequently classified as part of the baseline group, as illustrated in Figure 2.

[32]Specifically, we limit our sample to students from cities that straddle a province border, and a student can appear multiple times if their home city belongs to different city pairs. We control for city-pair-contest fixed effects (i.e., 3606 city-pair-contest dummies), and cluster the standard errors at both the province-subject level and city-pair-contest level.

[33]When the average quality of additional participants is lower in the presence of quota bonuses, assuming all other factors remain constant, we would anticipate a reduced probability of host-province participants winning a gold medal (conditioning on participation). This is especially relevant for CHSBO, where the host school is permitted to send its own team, which is likely to result in a lower average participant quality.

[34]Data on the ranking in Provincial Science Olympiads after 2006 is generally unavailable because (1) the name list of gold medalists in Provincial Science Olympiads is unavailable, and/or (2) the gold medalists are ranked by their surnames instead of performance.

0.4, suggesting that provincial rankings can be predictive of participant quality. Column (1) of Table A3 further exploits within-province-contest variation to corroborate this result. Moreover, students who ranked higher in the Provincial Olympiad are also more likely to win a gold medal in the National Olympiad (Column (2)). These findings suggest that additional participants, who ranked lower in the Provincial Olympiad compared to "regular" provincial team members, are more likely to perform poorly and have a reduced likelihood of earning a gold medal. Therefore, holding all else constant, sending additional participants should drive down the average probability of winning the gold medal for host-province participants. In other words, we should observe a negative coefficient for CPhO, CChO, CHSBO, and NOI in Table 3 in a world without cheating.

5.3.2 Overall Better Cohort?

Cohort sizes vary yearly, and a province might have more top students when its student population is larger. Meanwhile, it might be more inclined to apply as the host province for subjects when it expects to have more talented students. As a province generally would not host more than one subject in a year,, we have variation in the host province status across subjects within a province-year cell, allowing us to absorb temporal cohort shocks by adding province-year fixed effects in our baseline specification. Column (1) of Table 7 reports the result. The host province advantage does not disappear when the comparison was built on within-province-year variation. The finding remains robust when school-year fixed effects are used to exploit more stringent variation in Column (2).

In Column (4) of Table 7, we further adopt a different approach to test the "better cohort" explanation directly. If a host province has an overall superior cohort, this should translate into improved performance in other subjects, even if it does not host them. To this end, we introduce $HomeProv\_of\_OtherSubj$, a dummy that equals one if the participant is from a province hosting any academic Olympiad of other subjects in that year. We do not observe a host province extending its remarkable performance to other subjects that it does not host.

5.3.3 Subject-Specific Better Cohort?

Can province-year-subject quality changes explain the host province advantage? We discuss three prominent channels that merit consideration in this section. Furthermore, we leverage the centralized and stringent selection process for national team members, which is unrelated to college admission and beyond the control of host provinces, to demonstrate that our findings are not likely to be influenced by subject-specific quality improvement.

Strategical Reallocation. The host province may strategically reallocate talents from other subjects to the subject it hosts. Also, some talents may be attracted by the convenience/advantage of competing in their home province. However, this strategic reallocation is only theoretically possible. The National Science Olympiad features highly challenging questions, requiring participants to master undergraduate-level or even graduate-level knowledge. The host province status is announced one to two years before the event, and most participants begin subject-specific preparation in middle school, making rapid switches challenging. Even if reallocation were possible, it could deplete talent pools in other subjects, potentially lowering their performance. However, as shown in Column (3) of Table 7, a host province's strong performance in its designated subject does not negatively affect its performance in subjects it does not organize.

Infrastructure Investments. To achieve better performance, the host province could potentially invest heavily in subject-specific related infrastructure, such as teaching materials, instructor training, and facility updates. This can benefit both the current and future cohorts, appealing to local education authorities seeking a continuous stream of high-achieving students admitted to elite universities. In this case, we would expect a higher chance of winning gold medals in the years after hosting, which is especially true as the host province loses a quota bonus to send a more selective group. However, as shown in Figure 3, we do not find significant lagged effect from hosting.

Peer Effects. The host province often has more participants due to the quota bonus, and a larger pool of participants might facilitate the formation of study groups and collaboration, potentially enhancing learning outcomes. Column (4) of Table 7 investigates this possibility. The association between a province's number of participants and the probability of winning a medal is, however, negative. This suggests a higher number of participants may result in lower overall participant quality, possibly due to reduced resource allocation per individual or decreased collaboration stemming from increased competition.

Further Evidence from National Team Selection. China further screens top talents from gold medalists to select high schoolers to participate in the International Science Olympiads. In the first round, gold medalists in a certain upper quantile of the scores automatically advance into the list of *National Team Candidates*, who are eligible to move to the next round of selection. In the second round, the central committee (independent from the host province) organizes centralized training and additional rounds of tests to finalize the list of *National Team Members*.[35] After 2013, only *National Team Candidates* can secure admission to top universities (i.e., Tsinghua

[35] For example, there are 50 *National Team Candidates* for the International Mathematical Olympiad in 2021, while the final list of *National Team Members* consists of only 6 students.

University and Peking University) for certain, while other gold medalists need to take *Gaokao* but are generally rewarded with a large number of bonus points. As the second round of selection is not associated with college admission and beyond the control of host provinces, this setting offers a valuable opportunity for us to evaluate the quality of gold medalists from the host province.

Table 8 presents the results based on participant-level data ranging from 2013 to 2021.[36] Columns (1) and (2) show that, consistent with our baseline findings, participants from the host province are associated with a greater probability of winning a gold medal. Furthermore, they are even more likely to be ranked high enough to be automatically enrolled as *National Team Candidates* (and thus admitted by elite universities via the recommendation track). On average, the likelihood of being a *National Team Candidate* is 16.3% for participants in the National Olympiads; but this likelihood increases by 29.0% for students from the host province. Do *National Team Candidates* from the host province have better performance in the centralized selection of *National Team Members*? Column (3) shows that, within the *National Team Candidates*, students from the host province are associated with a 2.67 percentage-points decrease in the average likelihood of becoming a *National Team Member* (the p-value is 0.114). In other words, in the final round of selection, where the local education authority cannot intervene, host-province gold medalists experience a 30.9% reduction in the probability of winning, compared to other gold medalists from non-host provinces.

Together, the evidence from national team selection further strengthens the notion that the host province advantage is unlikely to be driven by subject-specific quality improvement. Instead, it appears to align more consistently with the rent-seeking behaviors of host provinces.

6 Discussion and Conclusion

This paper studies host favoritism in China's junior science Olympiads, a high-stakes selection platform for identifying talented youth. Combining province-contest and participant-level data, we show a substantial increase in gold medalists hailing from the host province, which has a direct association with elite college admission. The effect is particularly prominent in host provinces with higher levels of corruption and favor-exchange norms. We present evidence against alternative explanations such as comfort and improved quality. Instead, we document a range of suspicious patterns suggestive of local agents' cheating behavior – the unusual score premium observed in experimental parts, the absence of host-province effects

[36] The name list of *National Team Candidates* has become transparent and published online after 2013, as it has been directly related to college admission via the recommendation track since then.

once the experimental component is excluded, and the comparatively lower performance of host-province medalists in the centralized national team selection.

We believe our context is of special interest. The sample under our study is a group of (perhaps the most) talented high schoolers. Our results imply that fraudulent action even plays a non-trivial role in securing their admission into prestigious universities, raising concerns about the academic integrity of these future stars and its potential contagious effect (Ajzenman, 2021; Gulino and Masera, 2023).[37] Some of these individuals who later ascend to positions of power can wield significant influence over the generation of knowledge, encompassing the distribution of research resources and the trajectory of innovation (Fisman et al., 2018; Acemoglu, Yang and Zhou, 2022). More broadly, our results may provide a window into understanding the consequences of using decentralized non-standardized tests and associated signals to select top talent worldwide. A talent selection assessment, in contrast to the SAT or college entrance exams, is limited to a small number of participants possessing very specific abilities. The process involves customized and challenging questions designed by a small group of experts, while at the same time the opaque nature of the process limits social monitoring. It may therefore be necessary to promote transparency or involve independent oversight in order to ensure a fair selection process.

Apart from understanding biases in high-stakes talent selection, there are at least two noteworthy avenues.

First, while our empirical strategy holds constant the province-subject fixed effects to obtain a credible identification, as shown in Figure 2, the host provinces are unevenly concentrated in coastal and middle regions. That is, the host province status is indeed positively associated with local prosperity and educational resources. Accordingly, wealthier regions may have more bearing on favoring their own students and thus send more high schoolers to elite universities via the contest track, which can further undermine social mobility. This notion might be relevant to other contexts of high-stakes exams and talent selection, given that test sites (e.g., SAT and GRE) are usually located in better zones or institutions. Relatedly, the recent eye-opening "Singer's bribery scheme" case in the United States, in which wealthy families buy their children a place in top universities, also speaks to a similar notion – disparity self-propagate.[38]

[37] Given the relatively rigorous nature of Science Olympiads, our setting is believed to be one of the most challenging avenues for students seeking to unfairly gain admission into prestigious universities. However, the presence of cheating even in such scenario prompts compelling inquiries into the fairness of other decentralized talent selection systems (e.g., student-athletes), and should stimulate further investigation.

[38] See also the recent work of Chetty, Deming and Friedman (2023), which shows that U.S. Children from families

Second, we find that the unusual host-province advantage largely disappears after the removal of lab experiments, and there is no favoritism in the centralized national team selection. One may thus wonder why the Chinese government still favored the decentralized contest institution in the past decades. A potential rationale is that, as the ultimate goal is to select super talents to participate in the International Science Olympiads, decentralizing pre-selection (provincial/national Olympiads) with explicit and implicit bonuses can help mobilize local education authorities and students. As long as the central committee can screen the upper-tailed talents to form the national team in the final round (as shown by our analysis), the "host institution" – a common organizational approach for talent selection worldwide – seems to be a cost-effective instrument in this regard. Our findings thus suggest an analysis of the political economy nexus and an organizational view when considering policymakers' incentives to design and implement talent screening mechanisms.

in the top 1% are more than twice as likely to be admitted by an Ivy-Plus college as those from middle-class families with comparable SAT/ACT scores.

References

Acemoglu, Daron, David Y Yang, and Jie Zhou. 2022. "Power and the Direction of Research: Evidence from China's Academia." *Mimeo*.

Agarwal, Ruchir, and Patrick Gaule. 2020. "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights*, 2(4): 409–24.

Ajzenman, Nicolás. 2021. "The Power of Example: Corruption Spurs Corruption." *American Economic Journal: Applied Economics*, 13(2): 230–57.

Bai, Chong-en, Wei Chi, and Xiaoye Qian. 2014. "Do college entrance examination scores predict undergraduate GPAs? A tale of two universities." *China Economic Review*, 30: 632–647.

Borcan, Oana, Mikael Lindahl, and Andreea Mitrut. 2014. "The impact of an unexpected wage cut on corruption: Evidence from a "Xeroxed" exam." *Journal of Public Economics*, 120: 32–47.

Bulman, George. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics*, 7(4): 1–36.

Buser, Thomas, and Huaiping Yuan. 2019. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad." *American Economic Journal: Applied Economics*, 11(3): 225–52.

Chetty, Raj, David J Deming, and John N Friedman. 2023. "Diversifying Society's Leaders? The Causal Effects of Admission to Highly Selective Private Colleges." *Mimeo*.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." National Bureau of Economic Research Working Paper 23618.

Choudhury, Prithwiraj, Ina Ganguli, and Patrick Gaulé. 2023. "Top Talent, Elite Colleges, and Migration: Evidence from the Indian Institutes of Technology." *Journal of Development Economics*, 164: 103120.

Dee, Thomas S, Will Dobbie, Brian A Jacob, and Jonah Rockoff. 2019. "The causes and consequences of test score manipulation: Evidence from the New York regents examinations." *American Economic Journal: Applied Economics*, 11(3): 382–423.

Ellison, Glenn, and Ashley Swanson. 2016. "Do schools matter for high math achievement? Evidence from the American mathematics competitions." *American Economic Review*, 106(6): 1244–1277.

Ellison, Glenn, and Ashley Swanson. 2021. "Dynamics of the Gender Gap in High Math Achievement." *Journal of Human Resources*.

Ferraz, Claudio, Frederico Finan, and Diana B. Moreira. 2012. "Corrupting learning: Evidence from missing federal education funds in Brazil." *Journal of Public Economics*, 96(9): 712–726.

Fisman, Raymond, Jing Shi, Yongxiang Wang, and Rong Xu. 2018. "Social Ties and Favoritism in Chinese Science." *Journal of Political Economy*, 126(3): 1134–1171.

Goodman, Joshua, Oded Gurantz, and Jonathan Smith. 2020. "Take Two! SAT Retaking and College Enrollment Gaps." *American Economic Journal: Economic Policy*, 12(2): 115–58.

Gulino, Giorgio, and Federico Masera. 2023. "Contagious Dishonesty: Corruption Scandals and Supermarket Theft." *American Economic Journal: Applied Economics (Forthcoming)*.

Jacob, Brian A, and Steven D Levitt. 2003. "Rotten apples: An investigation of the prevalence and predictors of teacher cheating." *The Quarterly Journal of Economics*, 118(3): 843–877.

Lavy, Victor. 2009. "Performance pay and teachers' effort, productivity, and grading ethics." *American Economic Review*, 99(5): 1979–2011.

Moreira, Diana. 2019. "Success spills over: How awards affect winners' and peers' performance in Brazil." *Mimeo*.

Pettigrew, Stephen, and Danyel Reiche. 2016. "Hosting the Olympic Games: an overstated advantage in sports history." *The International Journal of the History of Sport*, 33(6-7): 635–647.

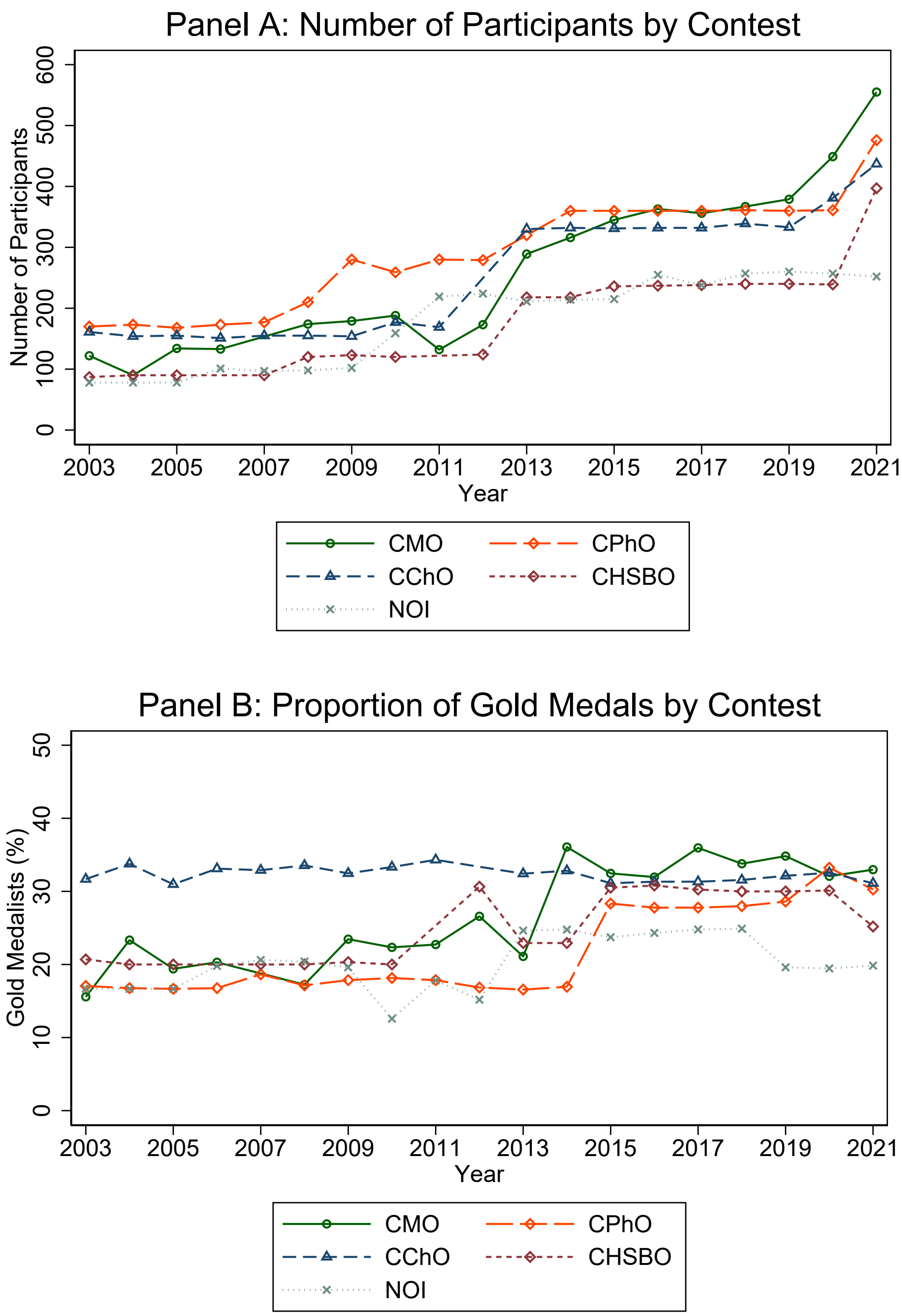
Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda*." *The Quarterly Journal of Economics*, 119(2): 679–705.

Reinikka, Ritva, and Jakob Svensson. 2005. "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda." *Journal of the European Economic Association*, 3(2-3): 259–267.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, 94(5): 1002–1037.

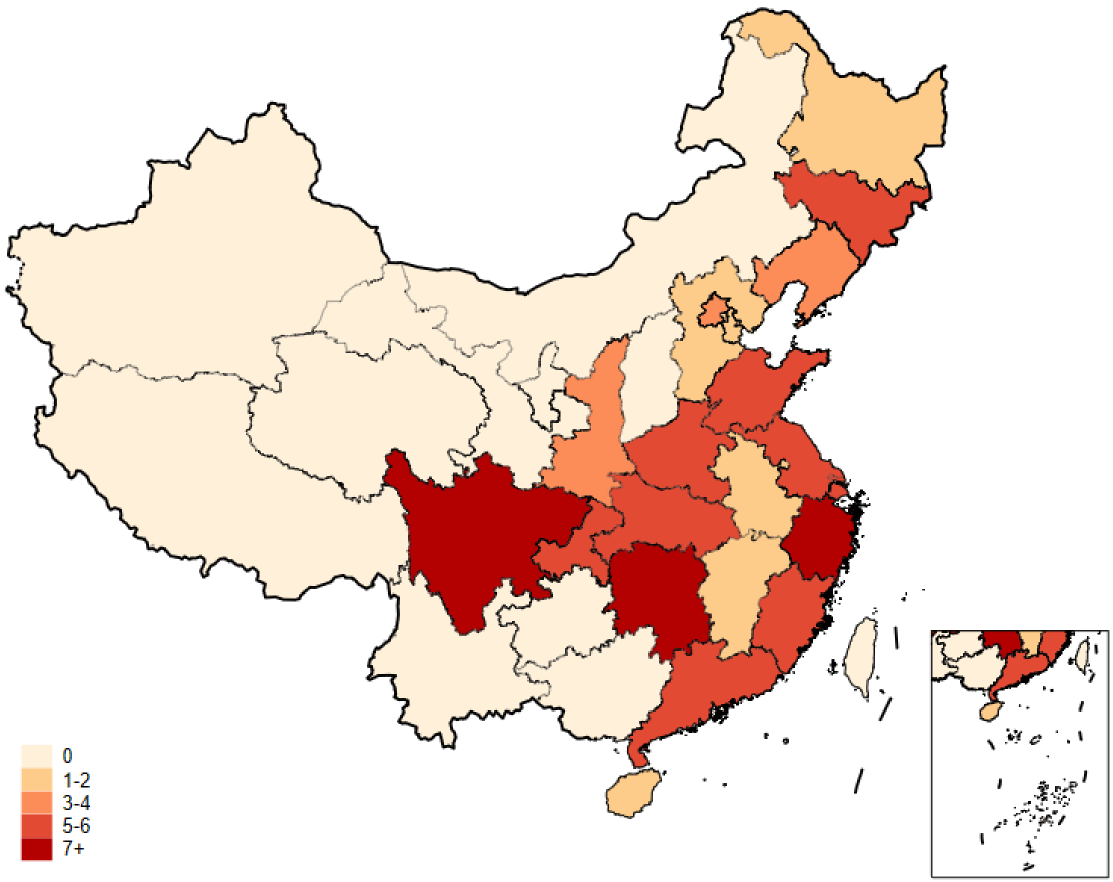
Figures

Figure 1: Participant Status by Contest



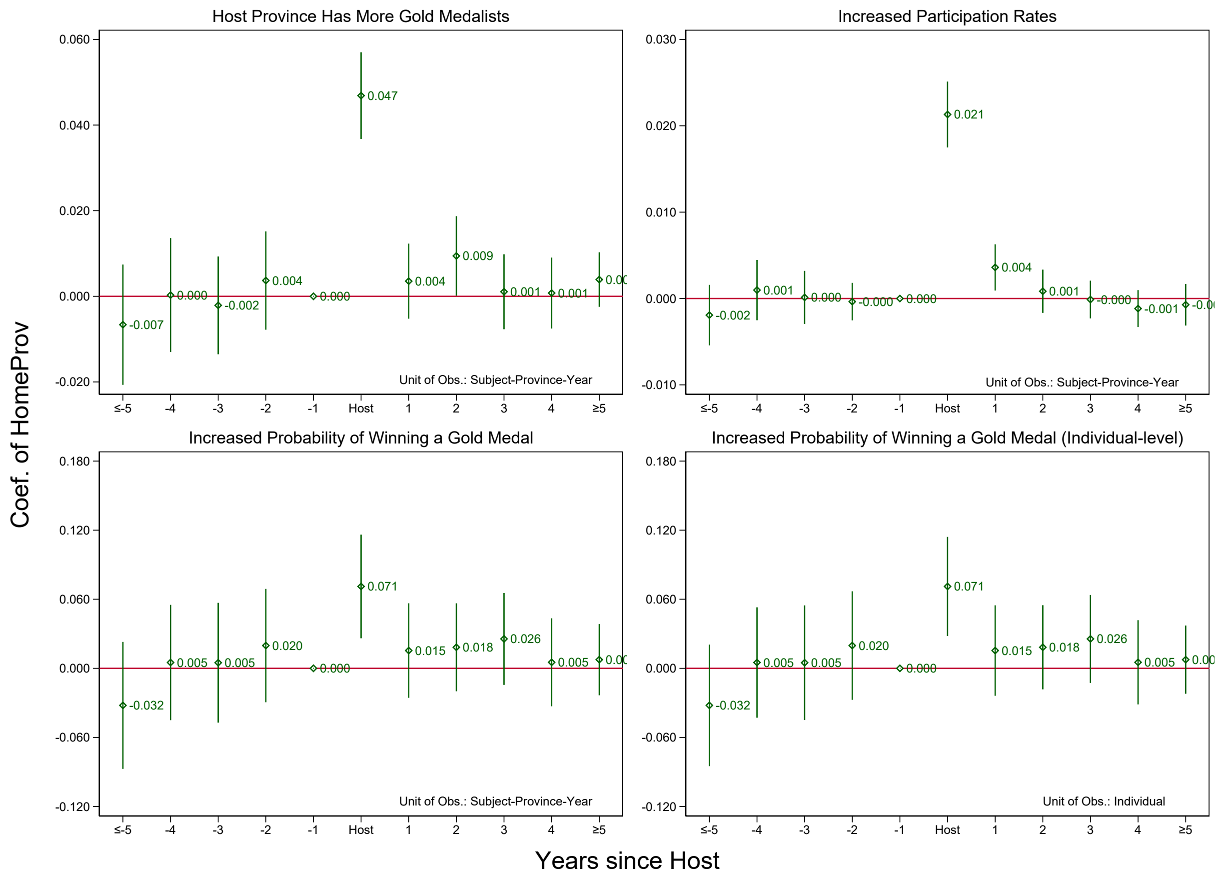
Notes: Panel A depicts the total number of participants in each Olympiad, and Panel B depicts the ratio of gold medalists in each Olympiad.

Figure 2: The Frequency of Being the Host Province (2003 - 2021)



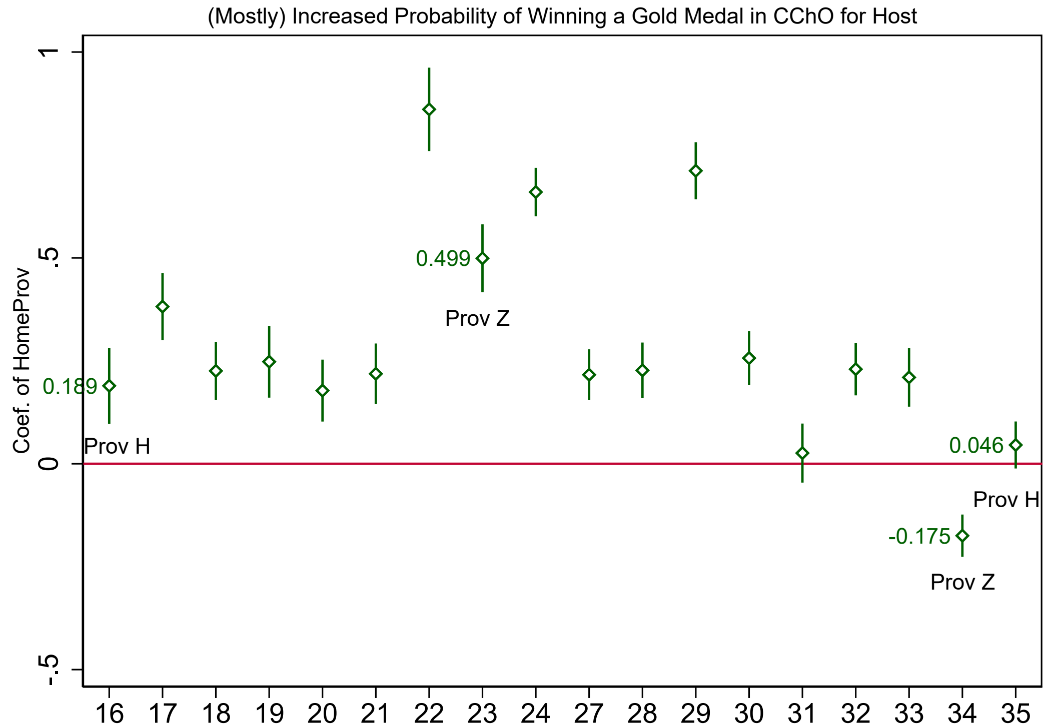
Notes: The figure depicts the number of times being the host for each province from 2003 to 2021 (pooling all five subjects).

Figure 3: Event Study Plots



Notes: The figure summarizes the results in Table 2, Table A1, and Table 3 using the event study specification. The (omitted) base period is 1 year prior to acting as the host province. Standard errors are clustered at the Province-Subject level.

Figure 4: Chemistry Contest Reform – Removal of Lab Experiments



Notes: The figure presents the dynamic impacts of acting as the host province for the Chinese Chemistry Olympiad (CChO) on the likelihood of winning a gold medal (conditional on participation). Unit of observation is the participant. CChO has removed the lab experiment part since its 34th contest.

Tables

Table 1: Summary Statistics

	Mean	S.D.	Min.	Max.	Obs.
Panel A. Province-Contest Level Variables					
Proportion of Gold Medals	0.033	0.05	0.00	0.31	2779
Proportion of Participants	0.033	0.02	0.00	0.12	2779
Prob. of being Gold Medalists	0.204	0.24	0.00	1.00	2653
# Gold Medals	2.069	3.30	0.00	35.00	2779
Number of Participants	7.788	5.75	0.00	41.00	2779
HomeProv	0.033	0.18	0.00	1.00	2779
Panel B. Individual Level Variables					
Gold Medalist	0.266	0.44	0.00	1.00	21643
HomeProv	0.065	0.25	0.00	1.00	21643
HomeCity	0.043	0.20	0.00	1.00	21643
SameProvOtherCity	0.022	0.15	0.00	1.00	21643
NeighboringProv	0.167	0.37	0.00	1.00	21643
Distance (100km)	10.608	6.44	0.00	40.54	21363
HomeProv_of_OtherSubj	0.147	0.35	0.00	1.00	21643
Standardized Score	0.000	1.00	-3.93	2.38	572
Lab Experinment	0.500	0.50	0.00	1.00	572
Ranking in Provincial Pre-Selection	2.948	1.67	1.00	10.00	1983
Ranking in National Olympiad	2.948	1.67	1.00	10.00	1983
National Team Candidate	0.163	0.37	0.00	1.00	14640
National Team Member	0.014	0.12	0.00	1.00	14640

Notes: The table presents the summary statistics of the main variables used in our analysis. The number of observations may vary due to the different time span and availability of the data, which we demonstrate in detail in the main text.

Table 2: Host Provinces Earn More Gold Medals

Panel A: Pooling all subjects	(1)	(2)	(3)	(4)
	Proportion of Gold Medals		# Gold Medals	
HomeProv	0.0493***	0.0457***	3.180***	2.549***
	(0.00515)	(0.00523)	(0.450)	(0.354)
Contest FEs	Y	Y	Y	Y
Province-Subject FEs	Y	Y	Y	Y
Province-Subject Time Trends		Y		Y
Observations	2,779	2,779	2,779	2,779
R-squared	0.674	0.740	0.657	0.834
Mean of DV	0.0331	0.0331	2.069	2.069
Cond. Mean of DV (Ever Host)	0.0442	0.0442	2.784	2.784
Cond. Mean of DV (Never Host)	0.00483	0.00483	0.274	0.274

Panel B: By subject	(1)	(2)	(3)	(4)	(5)
	Proportion of Gold Medals				
HomeProv	0.00507	0.0448***	0.0575***	0.0759***	0.0480***
	(0.00628)	(0.00912)	(0.00754)	(0.00978)	(0.0140)
Subject	Math	Physics	Chemistry	Biology	Informatics
Contest FEs	Y	Y	Y	Y	Y
Province FEs	Y	Y	Y	Y	Y
Province Time Trends	Y	Y	Y	Y	Y
Observations	589	589	540	510	551
R-squared	0.745	0.778	0.742	0.728	0.740
Mean of DV	0.0323	0.0323	0.0333	0.0333	0.0345
Cond. Mean of DV (Ever Host)	0.0403	0.0557	0.0495	0.0559	0.0680
Cond. Mean of DV (Never Host)	0.0250	0.0121	0.0154	0.0109	0.0100

Notes: Unit of observation is the Province-Subject-Year (Province-Contest). "Proportion of Gold Medals" refers to the number of gold medals a province earns in a given contest, normalized by the total number of gold medals in that contest. Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Non-decreasing Probability of Winning a Gold Medal – Participant-Level Evidence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prob. of being Gold Medalists	Gold Medalist (Binary)					
HomeProv	0.0638***	0.0638***	0.0223	0.0438	0.250***	0.0173	0.00724
	(0.0201)	(0.0192)	(0.0367)	(0.0349)	(0.0549)	(0.0410)	(0.0175)
Subject	All	All	Math	Physics	Chemistry	Biology	Informatics
Contest FEs	Y	Y	Y	Y	Y	Y	Y
Province-Subject FEs	Y	Y	Y	Y	Y	Y	Y
Prov-Subject Time Trends	Y	Y	Y	Y	Y	Y	Y
Observations	2,653	21,643	5,078	5,487	4,578	3,107	3,393
R-squared	0.699	0.202	0.145	0.176	0.254	0.295	0.127
Unit of Obs.	Province-Contest	Columns (2) - (7): Individual (i.e., conditional on participation)					
Mean of DV	0.266	0.266	0.295	0.230	0.322	0.263	0.207
Cond. Mean of DV (Ever Host)	0.306	0.306	0.298	0.305	0.414	0.419	0.263
Cond. Mean of DV (Never Host)	0.0593	0.0593	0.282	0.115	0.178	0.101	0.0948

Notes: In Column (1), "Prob. of being Gold Medalists" refers to the number of gold medals a province earns, normalized by the number of participants from that province in a given contest ($\frac{\text{\# Gold Medalists from Province P}}{\text{\# Participants from Province P}}$). As the denominator is at the province-contest level, Column (1) is weighted by the participation rate.Standard errors are clustered at the Province-Subject level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Heterogeneity by Host Province Corruptness

	(1)	(2)	(3)	(4)
	Gold Medalist			
HomeProv	0.0789***	0.0820***	0.0705***	0.0726***
	(0.0207)	(0.0219)	(0.0194)	(0.0204)
× Corruption norms (standardized)	0.0428**	0.0396*	0.0429**	0.0338*
	(0.0240)	(0.0233)	(0.0181)	(0.0203)
× Education spending per capita (standardized)		-0.0144		-0.0172
		(0.0163)		(0.018)
× GDP per capita (standardized)		-0.0192		-0.00414
		(0.0177)		(0.0214)
"Corruption Norms" Measure	CFPS survey: necessity of corruption		Search Index: gift money to teachers	
Contest FEs	Y	Y	Y	Y
Province-Subject FEs	Y	Y	Y	Y
Province-Subject Time Trends	Y	Y	Y	Y
Observations	21,643	21,643	21,643	21,643
R-squared	0.227	0.331	0.202	0.203
Mean of DV	0.266	0.266	0.266	0.266

Notes: Unit of observation is the individual (participant). "Corruption norms" is a province-specific measure. In Columns (1) and (2), we use the share of respondents agreeing with the survey question "for one to succeed, corruption is necessary", based on the Chinese Family Panel Studies (2010) – a nationally representative household survey. In Columns (3) and (4), we use the search index for the keyword "Money gift" (*hongbao* in Chinese) during Chinese Teachers' Day, normalized by the "Money gift" search index for the whole year (averaged between 2011 and 2021 when indexes are available). "Education spending per capita" and "GDP per capita" are at the province-year level; we include their first-order terms (direct effects) in our estimation but do not report them, which they are also not statistically significant. Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: Host Provinces and Lab Experiment Score Premium

	(1)	(2)
	Standardized Score (Individual-Level)	
HomeProv	0.157	0.0581
	(0.184)	(0.122)
Lab Experiment	-0.0130	-0.0551
	(0.0459)	(0.0797)
HomeProv × Lab Experiment	0.186***	0.338**
	(0.0615)	(0.129)
Subject	Biology (CHSBO 8, 16, 21)	
Contest FEs	Y	Y
Province FEs	Y	Y
Observations	572	150
R-squared	0.551	0.228
Sample	All	Gold Medalists
Mean of DV	0.000	1.045
Cond. Mean of DV (Ever Host)	0.662	0.985
Cond. Mean of DV (Never Host)	-0.125	1.021

Notes: Unit of observation is the individual (participant). "Lab Experiment" is a dummy that is 1 if the outcome corresponds to the test score of the lab experiment part, and 0 if the outcome corresponds to the test score of the theory part. We pool all contests where participants' theory and lab scores are available. Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Testing the Comfort Channel

	(1)	(2)	(3)
		Gold Medalist	
HomeProv	0.0546***		0.0721*
	(0.0200)		(0.0408)
Distance (100km)	-0.000985		
	(0.000683)		
HomeCity		0.0671***	
		(0.0237)	
SameProvOtherCity		0.0677**	
		(0.0290)	
NeighboringProv		0.0164*	
		(0.00893)	
Contest FEs	Y	Y	-
Province-Subject FEs	Y	Y	Y
Province-Subject Time Trends	Y	Y	Y
City-Pair-Contest FEs			Y
Observations	21,363	21,363	23,531
R-squared	0.203	0.203	0.314
Mean of DV	0.267	0.267	0.300
Cond. Mean of DV (Ever Host)	0.307	0.307	0.332
Cond. Mean of DV (Never Host)	0.0597	0.0597	0.229

Notes: Unit of observation is the individual (participant). "HomeCity" is a dummy that is 1 if the participant is from the host city of the contest; "SameProvOtherCity" is a dummy that is 1 if the participant is from other cities within the host province; "NeighboringProv" is a dummy that is 1 if the participant is from any neighboring province of the host province. Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Testing the Quality Channel

	(1)	(2)	(3)	(4)
	Gold Medalist			
HomeProv	0.0530***	0.0531**	0.0648***	0.0843***
	(0.0177)	(0.0232)	(0.0195)	(0.0208)
HomeProv_of_OtherSubj			0.00962	
			(0.0113)	
# of Participants				-0.00399**
				(0.00166)
Contest FEs	Y	Y	Y	Y
Province-Subject FEs	Y	Y	Y	Y
Province-Subject Time Trends	Y	Y	Y	Y
Province-Year FEs	Y	-	-	-
School-Year FEs		Y	-	-
Observations	21,642	19,327	21,643	21,643
R-squared	0.227	0.331	0.202	0.203
Mean of DV	0.266	0.280	0.266	0.266
Cond. Mean of DV (Ever Host)	0.306	0.320	0.306	0.306
Cond. Mean of DV (Never Host)	0.0593	0.0662	0.0593	0.0593

Notes: “HomeProv_of_OtherSubj” is a dummy that is 1 if the participant is from a province that hosts any academic Olympiad of other subjects in that year. Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8: Further Evidence from National Team Selection

	(1) Gold Medalist	(2) National Team Candidate	(3) National Team Member
HomeProv	0.0442*	0.0531***	-0.0267
	(0.0254)	(0.0203)	(0.0168)
Contest FEs	Y	Y	Y
Province-Subject FEs	Y	Y	Y
Prov-Subject Time Trends	Y	Y	Y
Sample	All	All	National Team Candidates
Observations	14,638	14,638	2,378
R-squared	0.209	0.145	0.067
Mean of DV	0.288	0.163	0.0845
Cond. Mean of DV (Ever Host)	0.329	0.183	0.0865
Cond. Mean of DV (Never Host)	0.0602	0.0277	0.0270

Notes: Unit of observation is the individual (participant). Each dependent variable is a binary indicating the corresponding outcome. *National Team Candidates* are selected from gold medalists based on their rankings in the contest (hosted by the host province), which is associated with elite college admissions. Among these candidates, China further selects its *National Team Members* who will attend international academic Olympiads; this additional round of selection is centralized and has no direct association with college admissions. Sample period: 2013-2021 (detailed information on national team members not available prior to 2013). Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix

Table A1: Host Provinces and Increased Participation Rates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Proportion of Participants	Number of Participants	Proportion of Participants (by Subject)				
HomeProv	0.0211***	4.420***	0.00550***	0.0258***	0.0102***	0.0334***	0.0318***
	(0.00180)	(0.415)	(0.00190)	(0.00170)	(0.000945)	(0.000685)	(0.00354)
Subject	All	All	Math	Physics	Chemistry	Biology	Informatics
Contest FEs	Y	Y	Y	Y	Y	Y	Y
Province-Subject FEs	Y	Y	Y	Y	Y	Y	Y
Province-Subject Trends	Y	Y	Y	Y	Y	Y	Y
Observations	2,779	2,779	589	589	540	510	551
R-squared	0.850	0.914	0.861	0.836	0.794	0.819	0.889
Mean of DV	0.0331	7.788	0.0323	0.0323	0.0333	0.0333	0.0345
Cond. Mean of DV (Ever Host)	0.0383	9.091	0.0402	0.0407	0.0377	0.0341	0.0532
Cond. Mean of DV (Never Host)	0.0202	4.614	0.0253	0.0244	0.0288	0.0308	0.0209

Notes: Unit of observation is the Province-Subject-Year (Province-Contest). "Proportion of Participants" refers to the number of participants from a province in a given contest, normalized by the total number of participants in that contest. Standard errors are clustered at the Province-Subject level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A2: Robustness to Alternative Clusters

	(1) Proportion of Gold Medals	(2) Proportion of Participants	(3) Prob. of being Gold Medalists	(4) Gold Medalist (Binary)
HomeProv	0.0457	0.0211	0.0638	0.0638
Cluster: Province-Subject	(0.00523)***	(0.00180)***	(0.0201)***	(0.0192)***
Cluster: Province-Year	(0.00523)***	(0.00138)***	(0.0184)***	(0.0171)***
Cluster (Two-way): Province & Year	(0.00408)***	(0.000917)***	(0.0202)***	(0.0197)***
Cluster (Wild-bootstrap P-value): Province	[<0.00100]***	[<0.00100]***	[0.00600]***	[0.00200]***
Contest FEs	Y	Y	Y	Y
Province-Subject FEs	Y	Y	Y	Y
Province-Subject Time Trends	Y	Y	Y	Y
Observations	2,779	2,779	2,779	21,643
R-squared	0.740	0.914	0.699	0.202
Unit of Obs.	Columns (1) - (3): Province-Contest			Individual

Notes: "Proportion of Gold Medals" refers to the number of gold medals a province earns in a given contest, normalized by the total number of gold medals in that contest. "Proportion of Participants" refers to the number of participants from a province in a given contest, normalized by the total number of participants in that contest. "Prob. of being Gold Medalists" refers to the number of gold medals a province earns, normalized by the number of participants from that province in a given contest. In parentheses, we show standard errors based on different clusters. In brackets, we report wild-bootstrap P values. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A3: Quota Bonus and Quality of Participants

	(1) Ranking in National Olympiad	(2) Gold Medalist
Ranking in Provincial Pre-Selection	0.203***	-0.0297***
	(0.0293)	(0.00682)
Province-Subject-Year FEs	Y	Y
Observations	1,957	1,957
R-squared	0.256	0.381
Sample Period	2003 - 2006	2003 - 2006
Mean of Dep. Var.	2.974	0.230

Notes: Unit of observation is the individual (participant). Smaller ranking indicate greater test scores (and thus performance in a given contest) – e.g., ranking=1 if the student has the highest score in the contest. Provincial Pre-selection: provincial team members (who attend National Olympiads) are selected based on their rankings in provincial pre-selection contests. The raw correlation between pre-selection rankings and rankings in national Olympiad is 0.4. Standard errors are clustered at the Province-Subject level. Detailed information on individual rankings are not available after 2006. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.