# Assessment Location and High-Stakes Cognitive Performance*

Justin Jihao Hong †  Victor Lei‡  Xuan Li§

This version: April 2026

**Abstract**

Merit is not directly observed but measured, and the conditions under which it is measured can shape how ability is revealed. This paper provides the first causal evidence that the location of assessments — a pervasive institutional friction in high-stakes testing worldwide — can substantially bias measured performance and the allocation of opportunity. Exploiting the random assignment of college entrance test centers in China, we find that students assigned to an off-site center score 0.14 standard deviations lower than classmates who test at their home school, reducing their likelihood of college admission by 3.8 percentage points. Mechanism analysis points to environmental unfamiliarity as a primary driver, while longer travel distances also play a role. We further document substantial equity implications: (1) effects are largest for disadvantaged students, and (2) the concentration of test centers in high-performing regions may inadvertently exacerbate these gaps. Finally, these measurement distortions have lasting economic consequences: affected students earn lower wages and are less geographically mobile.

**Keywords:** Assessment Location, Cognitive Performance, Measurement, Human Capital, Disparity, Education Infrastructure

**JEL:** D90, H75, I23, I24, I28

---

*We thank George Bulman, Yiming Cao, Raymond Fisman, Joshua Goodman, Kevin Lang, Daniele Paserman, Carolyn Pelnik, Eric Taylor, and participants at ACEM Doctoral Forum (SJTU, 2025), NSD Applied Micro Conference (PKU, 2025), NEUDC (Tufts, 2025), and PacDev (UCDavis, 2026) for their helpful comments. The data used in this study are accessed under strict supervision by the data provider and are fully de-identified. The authors have no financial or personal interests that are relevant to the results described in the paper. Each author contributes equally. All errors are our own.

†Hong Kong University of Science and Technology, Guangzhou (email: justinjihaoh@hkust-gz.edu.cn)

‡Harvard University and University of Hong Kong (email: ylei@law.harvard.edu)

§Boston University (corresponding, email: xuanli@bu.edu)

# 1 Introduction

From college admissions to employment screening, standardized assessments are ubiquitous (Hoekstra, 2009; Zimmerman, 2014, 2019; Anelli, 2020; Sekhri, 2020; Jia and Li, 2021). While these systems are designed to provide objective measures of human capital, the context in which ability is expressed can distort measurement (Gaggero and Tommasi, 2022). A pervasive global reality is that candidates — especially those from disadvantaged backgrounds — must often travel to unfamiliar locations to sit for exams that determine their future trajectories. In this paper, we present the first causal evidence that the location of high-stakes assessments can systematically bias measured performance and, in turn, shape the allocation of opportunity and longer-run economic well-being, with implications for both the accuracy and equity of human capital measurement.

Many high-stakes assessments are administered in off-site locations. In the United States, fewer than half of public high schools serve as SAT testing centers, a pattern especially pronounced in low-income communities; consequently, some students must travel hundreds of miles to reach an available center (Chiang, 2024). In many other contexts, organizers deliberately assign students to off-site centers to prevent cheating and facilitate administration — such as South Korea's CSAT, Brazil's ENEM, India's CBSE, and France's Baccalauréat. Despite its global significance, identifying the causal effect of assessment location remains challenging, because participation and test-site selection are typically endogenous. For instance, in the case of the SAT, students register voluntarily and pre-select their preferred test center from the available options. Conversely, in many mandatory assignment systems, useful variation is limited because all students are required to test off-site, and the assignment is not necessarily random.

In this paper, we overcome both challenges by exploiting a unique natural experiment: China's random assignment of students to test centers in the National College Entrance Examination (NCEE) — one of the world's most high-stakes assessments (Cai et al., 2019; Jia, Li and Cousineau, 2025). Nearly all high school graduates in China register for the NCEE, as university admissions are based solely on NCEE scores. After registration, local education authorities pool all NCEE candidates and *randomly* assign them to designated test centers (typically local high schools), rooms, and seats, with assignments fixed for the entire examination period. Crucially for our identification, this assignment mechanism generates *within-class* variation for students from high schools designated as test centers: While some remain at their own school to take the exam, others are required to take in different schools (see Figure A1 for a visual illustration). This setting enables comparisons between otherwise similar students who take the life-changing exam in different assessment locations.

Our empirical analysis leverages administrative data on the full population of approximately 11,000 students, who took the NCEE in a representative county between 2016 and 2018. The county's geographical size and population are comparable to those of Houston or Greater London. A unique advantage of our data, compared with the NCEE datasets used in existing studies, is that it contains detailed information on both students' assessment locations and their high schools of origin. It is also unusually rich in dimensions central to our identification and interpretation: We observe students' classes (strengthening identification), their home addresses (allowing precise computation of travel distance), and important downstream outcomes such as college admission, exam retake decisions, and individual labor market outcomes. Leveraging this unique dataset, we begin by validating the random assignment of test centers. Assignment to a non-home test center is not correlated with demographic characteristics, including gender, age, and socioeconomic status (SES).

Our main analysis estimates the impact of testing at a non-home school on exam performance. We find that students assigned to a non-home school score 0.14 standard deviations lower than those taking the exam at their home school — equivalent to a 10-point reduction on the 750-point scale used in China's NCEE. To put this magnitude in context, a gap of this size corresponds to roughly two additional incorrect multiple-choice questions (across all subjects). The non-home performance penalty persists throughout the sample period, and it is consistently observed across all high schools. In turn, these performance effects have tangible consequences for future opportunity: Students assigned to off-site test centers are ranked behind an additional 2.9% of their peers within the same year-track in their province, are 3.8 percentage points less likely to be admitted to any college, and are 0.6 percentage points less likely to gain admission to an elite college in the year of the exam. These consequences highlight that the context of high-stakes assessment can systematically affect the allocation of opportunity.

Longer-term labor market consequences further substantiate the significance of such assessment bias. Leveraging administrative tax records, we find that the initial location shock translates into real economic scarring. Students assigned to off-site test centers in their college entrance examination exhibit a reduction in early-career earnings of approximately 2 percent.

Furthermore, this educational setback significantly restricts their geographic mobility in the labor market, making them 3.5 percentage points less likely to work outside their home city. A back-of-the-envelope calculation suggests a benefit–cost ratio of approximately 28 from assigning all students to home-school test centers, relative to the current off-site system. Collectively, these findings demonstrate how cognitive distortions in life-changing assessments can alter individuals' economic trajectories.

We next examine the potential mechanisms, focusing on the two most relevant candidates: (1) testing environment and (2) travel distance. Two pieces of evidence suggest that the unfamiliar testing environment plays an important role. First, we find that having a familiar peer (i.e., a classmate) seated nearby mitigates the negative impact of testing in a non-home school, suggesting that peer familiarity helps recreate elements of a known environment and makes the setting feel less unfamiliar. Second, the performance penalty is concentrated in cognitive-demanding STEM subjects, consistent with the theory that unfamiliar settings interfere more with tasks requiring sustained concentration and complex reasoning (Beilock and Carr, 2005).

Turning to travel distance, we find that longer commutes are associated with slightly lower scores, even after controlling for socioeconomic background. This pattern aligns with the interpretation that longer travel may increase fatigue, reduce sleep time, or heighten uncertainty about reaching the test center on time. Nevertheless, controlling for travel time does not absorb the most estimated effect of testing at a non-home school.[1] We also carefully discuss other explanations such as cheating, and suggest they are not likely to be significant drivers in our context.

Finally, we discuss the broader implications for disparities. The performance penalty associated with testing location is both more severe for less privileged students and more likely to affect them. First, our *within-class* comparison shows that the penalty is concentrated among lower-performing and socioeconomically disadvantaged students, leading to worse admission outcomes for these groups. Although students can retake the exam in the following year, this option is primarily exercised by those from higher SES, likely due to the substantial financial and opportunity costs involved. Second, consistent with patterns observed in many other

[1] A plausible reason is that our within-class identification relies only on students from schools that serve as test centers (in order to leverage both home- and non-home testing variation). These test centers are geographically concentrated in urban areas. As a result, within our effective identification sample, students assigned to non-home schools travel only slightly farther on average than those testing at their own schools.

countries, test centers are disproportionately located in more developed areas and are therefore closer to more advantaged schools. As a result, students from less privileged backgrounds are more likely to be assigned to off-site test centers and thus disproportionately bear the performance penalty. A back-of-the-envelope calculation suggests that exam location accounts for about 7.6% of the observed performance gap between students from test-center schools and those from non-test-center schools.

This paper mainly contributes to the economics literature on educational inequality, by highlighting the role of bias in the measurement of human capital, distinct from its accumulation. A vast literature has examined how inputs and experiences — such as school resources, teacher effectiveness, family background, and early-life environment — shape human capital formation and perpetuate disparities (Krueger, 1999; Angrist and Lavy, 1999; Hoxby, 2000; Fredriksson, Öckert and Oosterbeek, 2012; Chetty, Friedman and Rockoff, 2014; Chetty, Hendren and Katz, 2016; Jackson and Mackevicius, 2024; Biasi, Lafortune and Schönholzer, 2025). In contrast, much less attention has been devoted to the features of *how* human capital is measured, including what is tested (Landaud et al., 2024; Lee and Schaelling, 2024; Li et al., 2024; Franco and Povea, 2025) and how it is administered (such as the choice of timing and location). Yet these factors may systemically influence how human capital is expressed and, consequently, how future opportunities are allocated.

A notable exception is Gaggero and Tommasi (2022), who use data from a UK public university and exploit the quasi-random assignment of exam times (9:00 a.m., 1:30 p.m., and 4:30 p.m.). They find that students tested in the early afternoon score 0.07 standard deviations higher than those tested earlier in the day, consistent with fluctuations in circadian rhythms. We advance this agenda by isolating the causal impact of another key dimension — assessment location — and linking it to real socioeconomic outcomes and implications for inequality. Leveraging China as a testing laboratory, we find that students randomly assigned to non-home sites suffer a performance penalty of 0.14 standard deviations — a magnitude comparable to the timing effect documented by Gaggero and Tommasi (2022). Our data further allow us to document real consequences: affected students are 3.8 percentage points less likely to gain college admission, leading to an average starting-wage penalty of 2%. Moreover, there exist notable implications for equity: this "location penalty" is most severe for low-SES and low-achieving students, who are least able to compensate for these setbacks through costly retaking.

Second, our study also relates to research on unequal access to educational resources, and more specifically to inequality in assessment infrastructure (Bulman, 2015; Smith, 2018). In the United States, fewer than half of public high schools serve as SAT test centers, with even lower rates among schools serving low-income communities. Bulman (2015) find that opening a new test center increases test-taking by an average of 8.5 percent among students at the host school, with roughly 40 percent of these additional test-takers subsequently enrolling in a four-year college. We make progress by shifting the focus from the extensive margin (whether to take the assessment) to the intensive margin (how assessment location affects performance conditional on participation). The "double penalty" associated with assessment location, operating through both access and biased performance, illustrates how the seemingly neutral design of assessment systems can reinforce inequities. Given the uneven distribution of test centers and prevalence of off-site testing in most countries, our findings may offer additional insights into other contexts.

Finally, we complement the growing literature on how transitory external shocks — such as pollution, temperature fluctuations, pollen exposure, neighborhood crime, the SNAP benefit cycle, and other exogenous disruptions — affect cognitive performance (Bensnes, 2016; Ebenstein, Lavy and Roth, 2016; Graff Zivin et al., 2020; Park, 2022; Bond et al., 2022; Chang and Padilla-Romo, 2023; Wang, Wang and Ye, 2023; Huang et al., 2025). Similar in spirit to these studies, we provide new evidence on how external conditions shape evaluation outcomes, but with a critical distinction regarding policy tractability. While environmental and socioeconomic shocks are often beyond policymakers' immediate control and costly to mitigate, institutional design elements — such as where exams are administered — are policy-leveraged and straightforward to adjust. Our back-of-the-envelope calculation implies that assigning all students to home-school test centers would yield a benefit–cost ratio of approximately 28, relative to the current off-site system.

# 2 Background and Data

## 2.1 Setting: National College Entrance Examination (NCEE)

In China, the National College Entrance Examination (NCEE), or *Gaokao*, is widely regarded as one of the most demanding and consequential standardized tests in the world. It functions as the primary — and often the sole — criterion for university admission. The exam is administered only once a year, in early June, and spans two consecutive days. For most students, this single exam largely determines their access to higher education and profoundly influences their long-term career prospects (Jia and Li, 2021).

**The Organization of NCEE (*Gaokao*).** The Chinese government places strong emphasis on the organization of the NCEE to ensure a smooth and efficient process. Its registration period takes place from November 1 to November 10 of the preceding year. In each county, several high schools are designated as test centers. All registered students within the same county take the exam at these centers, which typically contain 20 to 40 testing rooms, each accommodating 30 students.

Since 2016, as part of national efforts to further standardize the college entrance examination, China's Ministry of Education has required local authorities to randomly assign students to test rooms within designated administrative zones. The randomization is based solely on academic track (science or liberal arts), ensuring that students in a given room follow the same track. However, provinces differ in whether students may take the exam in their own schools, and our context has advantages in this regard. In some provinces, all or most students test in their home schools if those schools are designated as test centers. In others, including our sample province, assignment is fully random: students are allocated with equal probability to any test center within each city or county. Consequently, among students from high schools that serve as test centers, some are randomly assigned to take the assessment at their home school, while others are randomly assigned to different test centers.

The provincial government randomly assigns students to specific test centers, rooms, and seats one week before the exam.[2] Students are notified of their assigned test center, room, and seat several days after the randomization, and these assignments remain fixed throughout the entire examination period. Switching test centers, rooms, or seats is strictly prohibited. Every seat in the exam room is occupied, and proctors strictly enforce seating arrangements. Each assigned seat is labeled with the student's name, photo, and identification details, which are verified by proctors before the start of each exam.

Students are allowed to visit the test center the afternoon before the exam to familiarize themselves with their assigned room and nearby facilities. The test rooms will be re-inspected after the students' familiarization visit. Participation in this pre-exam visit is optional. During the examination, each room is supervised by two proctors — one male and one female — who

[2] A computer program is employed to randomly assign students to test rooms. Specifically, within each county and academic track, students are allocated according to computer-generated random numbers.

are high school teachers recruited from other counties and randomly assigned. After the exam, the papers are scanned by computer and then randomly assigned to experienced graders for evaluation in a double-blind process. Appointed by the provincial education authority, these graders convene for one week at a government-designated location to evaluate exam scripts.

Notably, since the NCEE is a high-stakes exam that students spend 12 years preparing for, we find that in our county, all registered students ultimately sit for the exam. In other words, whether students take the exam is unrelated to their assigned assessment locations in our context.

**Exam Subjects.** During our sample period (2016 – 2018), our sample province adopted the National Unified Exam Paper, which is designed and administered by the National Education Examinations Authority under the Ministry of Education. This exam was used in most provinces across China, meaning that a large number of students nationwide faced the same exam structure, content, and scoring standards. The exam consists of four sections totaling 750 points: Chinese (150), Mathematics (150), English (150), and a track-specific comprehensive test (300). There are generally two tracks: the science track and liberal arts track. The comprehensive test covers physics, chemistry, and biology for science track students, and politics, history, and geography for liberal arts track students. The selection of academic track occurs at the end of Grade 10 (two years before the NCEE). Students could not choose individual subjects within their track, as subject combinations are fixed.

**College Admission.** Students typically receive their test scores about two weeks after the exam. Upon receiving their scores, they begin formulating their college application lists. Notably, college admissions are administered separately by province, with each province setting its own admission quotas, cut-off scores, and ranking systems. As a result, students compete against peers within their own province.

Admission outcomes depend on both the number of available spots and a student's rank among provincial peers applying to the same institution. While approximately 40% gain admission to college, only a small fraction are accepted into top-tier universities, resulting in an intensely competitive environment with extraordinarily high stakes. Given the within-province competition, the NCEE is highly competitive — where even a single point can significantly affect a student's rank. To illustrate this competitiveness, consider our sample province in 2018, which had approximately 200,000 science track examinees. In this context, a student with a score of 500 achieved a provincial rank of 63,000. A reduction of just one point would lead to a drop of 680 positions, disadvantaging the student relative to hundreds of peers in the college selection process.

## 2.2 Data

Our study uses two primary data sources. The first is student-level examination data from the Bureau of Education of a county in Central China. The second is long-term individual tax records obtained from the State Taxation Administration. This administrative examination dataset covers all students who registered for the national college entrance examination in our sample county from 2016 to 2018. It contains detailed information on each student's gender, ethnicity, age, academic track, high school and class, test center, and test room. The dataset also includes subject-level test scores and, importantly, college admission outcomes. While there is a growing body of studies using different sources of NCEE data, to our knowledge, ours is the only one that contains granular information on specific assessment locations linked to individual student performance. Furthermore, to examine longer-term consequences, we obtain individual tax records for these students from the State Taxation Administration. Using these records, we construct two measures of labor market outcomes: starting earnings after graduation and the city of employment. Our tax records extend through 2025 and capture individuals' earnings reported in tax filings, allowing us to observe labor market outcomes for nearly all students in our sample.[3]

The county has a geographical size comparable to that of Houston or Greater London. Table A1 provides background information regarding the sampled county characteristics. For comparison, we also provide corresponding statistics for an average county. While our sampled county is less prosperous than the average, it has a comparable share of high schoolers in its population and a similar student-teacher ratio relative to other counties in the same province.

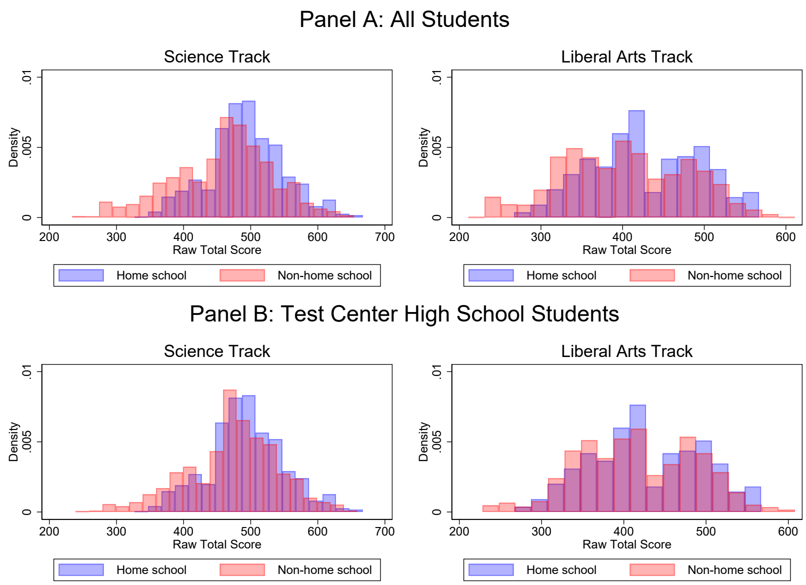
Each year, approximately 4,000 students from eleven high schools in our sample county register for the NCEE. Five of these schools serve as designated test centers, each accommodating a roughly equal number of students.[4] Specifically, the largest test center hosts 21% of students, while the smallest hosts 18.4%. Figure A2 presents the spatial distribution of high

[3] All procedures involving tax data were conducted under strict anonymization and confidentiality protocols. Data linkage and variable construction were performed by authorized staff at the State Taxation Administration within a secure data facility, and researchers accessed only anonymized outputs.

[4] These five high schools are the top-performing schools in the area. We provide a detailed comparison of test-center and non–test-center schools in Section 4.3.

schools and designated test centers to show their relative positions. All five test centers are centrally located within the county situated closer to the local government offices. Students from the most remote high school must travel 9 kilometers — often navigating mountainous roads by bus or electric bicycle — to reach their designated test center for the exam.

**Figure 1: Distribution of Raw Total Scores**




*Notes:* This figure shows the distribution of raw total scores by test center status (home school vs. non–home school), separately for students in the science and liberal arts tracks. Panel A includes all students, while Panel B focuses on students attending high schools that serve as test centers. The histograms are based on raw data. Blue bars represent students who took the exam at their home schools, while red bars represent those who took the exam at non-home schools. The x-axis displays raw total scores, and the y-axis indicates the density.

Figure 1 presents the distribution of raw total scores for students in two academic tracks, comparing those who took the exam at their home school (in blue) *versus* a non-home school (in red). The left figure in Panel A shows the distribution for the Science Track. Both groups exhibit approximately bell-shaped distributions, but home-school test takers tend to cluster at slightly higher score ranges. On average, students who took the exam at their home school scored 492 points, compared to 459 points for those who tested at a non-home school, indicating a substantial performance gap. The right figure in Panel A shows the distribution for the Liberal Arts Track, where the score distributions are more similar across test center types. Nonetheless, home-school test takers still show a slight concentration in the upper score range, with average scores of 427 for home-school test takers and 401 for their non-home school counterparts. Overall, the figure highlights consistent performance differences by test center type: students who take the exam at their home schools tend to perform better. However, this initial comparison reflects the influence of two factors: (1) the potential performance bias driven by assessment locations, and (2) underlying differences between students from schools that serve as test centers and those from schools that do not (all of whom take the exam in non–home centers).

Because we focus on isolating the effect of assessment location on cognitive performance, our analysis centers on students from the five high schools in the county that serve as test centers. These schools offer meaningful *within-school* and *within-class* variation in assessment location, as each includes some students testing at their home school and others to one of several off-site centers. In contrast, students from schools that are not designated as test centers must take the exam elsewhere and thus exhibit no variation in home-school testing status. These five top high schools enroll 81% of all students in the county. Among them, 82.9% took the exam at a non-home school. The maximum distance between any two of these schools is 6.5 kilometers. Panel B of Figure 1 presents the distribution of raw total scores for students from the five high schools that serve as test centers, disaggregated by the location where they took the exam. While the score gap between home-school and non-home-school exam takers has narrowed, students who took the exam at their home school still clearly performed better.

A potential concern with relying on administrative data from a single region is external validity. We address this by highlighting four key features of our setting that support the broader applicability of our findings. First, as noted, the sampled county is demographically representative, capturing the educational environment faced by a typical student in China's hinterland. Second, the institutional environment is highly representative. The random assignment of test centers in our sample is not an idiosyncratic local policy but a standard implementation of China's national NCEE regulations. The frictions we study echo the challenges faced by millions of students across the country every year. Third, the reliance on a single county is necessitated by exceptional data scarcity. While the practice of randomization is widespread, detailed records of test administration are typically maintained only by local county-level education bureaus, rather than centralized provincial or national databases. Consequently, administrative data linking students' home addresses, their assigned test centers, and their exam scores is largely inaccessible due to both decentralized data management and strict privacy regulations.[5] Our dataset thus provides a window into the "black box" of a nationwide phenomenon. Fourth, the mechanisms driving our results, as discussed later, reflect fundamental cognitive and behavioral responses to unfamiliar environments and travel distance, suggesting that the "location penalty" is likely to arise in high-stakes assessments globally.

# 3 Empirical Strategy

## 3.1 Identification

In this section, we move beyond descriptive evidence and present more rigorous analysis. Our baseline specification is defined as follows:

$$TotalScore_{ic(s)r(t)} = \beta \times NonHomeSchool_{ic(s)r(t)} + \pi_{c(s)} + \pi_{r(t)} + \epsilon_{ic(s)r(t)} \tag{1}$$

where student $i$ attends high school $s$, belongs to high school class $c$, and takes the NCEE in test room $r$ at test center $t$. $TotalScore_{ic(s)r(t)}$ represents student $i$'s standardized total score on the NCEE. To ensure comparability of test scores across different years, we standardize test scores within each year and academic track (i.e., liberal arts or science) to have a mean of 0 and a standard deviation of 1. $NonHomeSchool_{ic(s)r(t)}$ is a dummy variable that is one if the student $i$ is assigned to a test center outside their own home school (i.e., $s \neq t$).

Crucially for our identification, we introduce $\pi_c$, the high school class fixed effects, which allow us to compare students within the same class and thus absorb any *high school* $\times$ *year* $\times$ *track* differences. In China, students are typically placed into classes according to their academic track and ability at the outset of 11th grade. Each class consists of 25 to 45 students with similar academic performance, and these students attend all of their courses together with the same group of classmates. The inclusion of high school class fixed effects enhances the comparability by controlling for these class-level groupings. This approach also addresses potential selection concerns, as some schools host a larger number of exam takers, and students from these schools are more likely to take the exam at their own institution. Given the inclusion of high school class fixed effects, our analysis focuses on students from the five high schools

[5] Similar to the restricted access protocols for high-granularity administrative data in other contexts (e.g., Brazil's ENEM), our data access required physical presence at the local government's secure terminals.

that serve as test centers (approximately 8,500 students).[6]

We also include test room fixed effects, $\pi_r$, which account for differences in the testing environment among students assigned to the same room, capturing factors such as temperature and the condition of school facilities. $\beta$ is the coefficient of interest, capturing the performance gap attributable to assignment to non-home test centers. Standard errors are clustered at the high school class level.

## 3.2 Balance checks

Given the random assignment process, it is reasonable to expect that a student's assignment to a non-home test center is uncorrelated with individual-level characteristics. To validate the randomness of test center assignment, we examine whether a student's placement at a non-home center is associated with individual-level characteristics.

Table A2 presents summary statistics and balance checks for demographic and background variables available in our data, comparing students taking exams at home schools and non-home schools. 49.3% of students are male, the average age is 17.6 years, 38.6% of students originate from urban areas, 81.7% are enrolled in the science track, and 52.3% live in neighborhoods with house prices above the median (i.e., ¥2500 per square meter).[7] The results show no statistically significant differences between the two groups in terms of gender, age, class leadership roles, participation in the science track, or socioeconomic status (proxied by the housing prices). Moreover, approximately 22% of students have a classmate in the same test room, and this proportion is identical for those testing in their home school and in other schools, as all students are randomly assigned to test centers and test rooms. These findings corroborate successful randomization and comparable student populations in our context.

[6] It is important to note that all students in our analysis are first-time exam takers, as these five high schools do not admit repeat test-takers, who are typically enrolled in specialized repeat-year schools. In Section 4.1, we provide additional analysis of repeat exam takers who graduated from these five schools in the previous year and find similar results.

[7] Notably, the urban–rural classification is based on household registration and thus reflects students' places of origin. The house price indicator is based on current home addresses and serves as a better proxy for socioeconomic status.

# 4 Results

## 4.1 Main results

Table 1 presents our main results. We begin by regressing test scores on an indicator for taking the assessment at a non-home center, including only high school class fixed effects. This approach enables comparisons among students within the same high school class — a comparable group with similar academic backgrounds and an equal likelihood of being assigned to different test centers. We find that students who take the exam at their home school perform substantially better. In Column (2), we include test center fixed effects to account for factors specific to each center, such as the condition of facilities. All high schools that served as test centers in our sample county underwent renovations between 2006 and 2008 and thus have relatively modern facilities. We again find that students who take the exam at their home school perform substantially better, with a similar magnitude of effect.

**Table 1: Main Results**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total Score | | | | Rank Percentile | College Admission | Elite College Admission | Exam Retake | Log. Earnings | Migration |
| Non-Home School | -0.136*** | -0.133*** | -0.141*** | -0.140*** | -2.930*** | -0.0375*** | -0.00583* | 0.133*** | -0.0195*** | -0.035*** |
| | (0.0174) | (0.0172) | (0.0186) | (0.0184) | (0.425) | (0.0103) | (0.00330) | (0.0115) | (0.0044) | (0.0114) |
| Observations | 8,535 | 8,535 | 8,535 | 8,535 | 8,535 | 8,535 | 8,535 | 5,619 | 8,535 | 8,535 |
| R-squared | 0.486 | 0.487 | 0.512 | 0.520 | 0.550 | 0.465 | 0.215 | 0.100 | 0.278 | 0.350 |
| Individual Controls | . | . | . | X | . | . | . | . | . | . |
| Test Center FEs | . | X | . | . | . | . | . | . | . | . |
| Test Room FEs | . | . | X | X | X | X | X | X | X | X |
| Highschool Class FEs | X | X | X | X | X | X | X | X | X | X |

*Notes:* This table presents our main results. $Non\text{-}Home\ School$ is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. $Total\ Score$ represents the student's total test score, standardized by year and academic track. $Rank\ Percentile$ measures provincial rank percentile. $College\ Admission$ is a binary variable indicating whether student $i$ is admitted to any college in that year. $Elite\ College\ Admission$ is a binary variable indicating whether student $i$ is admitted to an elite college in the current year. $Exam\ Retake$ is a binary variable indicating whether student $i$ retook the exam in the following year. Because our data span the 2016–2018 cohorts, we can identify retakers only for the 2016 and 2017 cohorts. Accordingly, Column (8) restricts the analysis to these two cohorts. $Log(Earnings)$ measures the log of starting earnings after graduation. $Migration$ is a binary variable indicating whether the student's workplace is located outside their home city. Individual controls include gender, age, class leadership status, urban residency, the presence of classmates, and socioeconomic status, proxied by the housing prices of students' current residential addresses. Standard errors are clustered at the high school class level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Column (3) presents our preferred specification, which incorporates both high school class fixed effects and test room fixed effects. The inclusion of test room fixed effects accounts for test room-specific factors, such as floor location, student composition, proctors, and other environmental influences. The estimated coefficient remains virtually unchanged. In Column (4), we add individual controls, including gender, age, class leadership status, urban residency, the presence of classmate in the same test room, and socioeconomic status. The results remain consistent. On average, students assigned to a non-home school score 0.14 standard deviations lower than their classmates who take the exam at their home school.

In addition, Column (1) of Panel A in Table A3 reports an alternative measure of exam performance: the raw total score. Students who take the exam at a non-home school score, on average, 10 points lower (out of 750), representing a 1.3% decrease in the total score. To put this into perspective, since each multiple-choice question typically carries 4 to 6 points, this gap is roughly equivalent to answering two more questions incorrectly across all six subjects.[8]

The importance of score changes is due to their influence on college attendance outcomes. Upon receiving their NCEE scores, students begin formulating their college application lists. Chinese universities are grouped into different admission tiers — First, Second, and Third Tiers — based on their overall quality or prestige. In each tier, students are allowed to apply to up to five schools. Universities admit students by selecting the highest-scoring applicants who listed them as a preference, continuing until their admission quotas are filled. Once a student is admitted by a university, they are removed from the applicant pool. As a result, each student can be admitted to only one university. If their score falls short for all the universities they applied to, they will not receive an offer of admission.[9]

Since college admission outcomes in China depend on a student's relative ranking among applicants from the same province and academic track (rather than on their absolute test scores), it is important to understand how changes in test scores translate to shifts in provincial ranking percentiles. In Column (5) of Table 1, we match our data to the score distribution in the province, accounting for year and academic track, and calculate the provincial rank percentiles, which measures the proportion of students with lower test scores within the same year-track. On average, the decline in test scores places affected students behind an additional

[8]To understand how this magnitude compares to the effects of short-run environmental and psychological shocks on student performance in other contexts, consider the following examples: in Mexico, violent crimes occurring in the week prior to exams reduce female students' test scores by 0.11 standard deviations (Chang and Padilla-Romo, 2023); in Israel, a 10-unit increase in PM2.5 exposure lowers scores by 0.08 standard deviations (Ebenstein, Lavy and Roth, 2016). Moreover, Park (2022) finds that taking the Regents Exams in New York City when outdoor temperatures reach 90°F reduces performance by approximately 0.13 standard deviations compared to taking the exam at 75°F. In the Chinese context, a 5°C (9°F) increase in temperature during the national college entrance exam period reduces total test scores by 0.15 standard deviations (Graff Zivin et al., 2020). Students taking the NCEE in China may be more sensitive to external environmental shocks due to its exceptionally challenging nature, coupled with the intense competition and high pressure surrounding it.

[9]For any given student, typically only one of the three tiers is relevant, depending on their exam score. For example, a high-scoring student will be admitted to a university in the First Tier, making any of their applications to schools in the Third Tier effectively irrelevant.

2.9% of their peers within the same year-track in the province.

Subsequently, in Column (6) of Table 1, we find that the decline in test scores translates into a meaningful reduction in immediate access to educational opportunities. In our sample county, 67.6% of students are admitted to college in the year of their exam. However, those assigned to non-home test centers are 3.8 percentage points less likely to be admitted (about 5.6% decrease). Among more than 2,000 universities in China, 39 are designated as top-tier institutions under the "985 Initiative". Column (7) shows that students assigned to non-home schools are 0.6 percentage points less likely to gain admission to one of these elite universities in the year of their exam. It is worth noting that this result should be interpreted with caution though, as admission outcomes are influenced by both test scores and students' application strategies (Li and Qiu, 2023).

Furthermore, NCEE takers who believe their initial scores do not accurately reflect their true abilities or meet their expectations may choose to retake the exam the following year, even if some have already been admitted to a college in the current year (Kang et al., 2024). These retakers typically remain in their home county to prepare for and sit the exam again. Because our data cover the population of exam takers in the county from 2016 to 2018, we can identify retakers only for the 2016 and 2017 cohorts. Accordingly, we restrict our analysis to these two cohorts. Specifically, we match observations across two consecutive years using full name, exact date of birth, gender, and academic track (science or liberal arts). Individuals successfully matched to records in the subsequent year are classified as having retaken the NCEE. In our sample county, approximately 19.7% of students retook the NCEE in the following year. In the last column of Table 1, we find that students assigned to a non-home test center are significantly more likely to be dissatisfied with their admission outcomes and are 13.3 percentage points more likely to retake the exam the following year.

**Longer-term Labor Outcomes.** After examining short-run educational outcomes, we complement with longer-term labor market consequences to substantiate the economic significance of the assessment bias. Using individual-level administrative tax records from the State Taxation Administration, Columns (9) and (10) of Table 1 examine early-career earnings and geographic mobility after graduation. Column (9) shows that students assigned to a non-home test center earn significantly lower starting earnings after entering the labor market. The estimated coefficient implies that these students earn approximately 2 percent less than comparable students who took the exam at their home school. Column (10) further examines geographic mobility by measuring whether an individual's workplace is located outside their home city. We find that students assigned to non-home test centers are 3.5 percentage points less likely to work outside their home city, relative to a baseline mobility rate of 51.9 percent in our sample.

These patterns suggest that the documented distortion in measured performance may have persistent effects on individuals' economic opportunities and spatial mobility. A back-of-the-envelope benefit–cost analysis in Appendix C indicates that the benefit–cost ratio of a policy assigning all students to home-school test centers (relative to the current system, in which most students are assigned to off-site venues) is approximately 28.2.

**Robustness.** We conduct an array of additional analyses to better characterize the non-home exam-taking penalty. We use alternative clustering levels in Panel B of Table A3. In Column (1), standard errors are clustered at the school-year-track level, while in Column (2), they are clustered at the test room level.

In Column (3) of Panel B in Table A3, we present additional analysis of repeat exam takers who graduated from five high schools serving as test centers. A student can retake the exam the following year if they are not satisfied with their first attempt, and some may be randomly assigned to their high school alma mater for their second attempt. Because of limited within-class variation in this specification — resulting from the small number of exam retakers — we instead include only school-track-year fixed effects.[10] Additionally, we control for their test scores from their previous NCEE, which serve as a baseline measure of academic ability in a high-stakes examination setting. We find similar results: these students perform better when taking the exam at their *alma mater*.

Table A4 presents heterogeneity results disaggregated by year, gender, and high schools. In Panel A, we demonstrate that our results are not driven by a specific year, as the effect persists throughout the sample period. Furthermore, this performance penalty is gender-neutral, with both male and female students experiencing a similar effect. Panel B of Table A4 further reports results disaggregated by high school. The observed effect is not driven by a particular school; rather, it is consistently present among students from all high schools that serve as

[10] Repeat test takers cannot enroll in regular high schools and instead attend specialized repeat-year schools. In this specification, school-year fixed effects refer to their original high schools rather than the specialized repeat-year schools.

test centers.[11] Although the magnitude varies across schools, this likely reflects differences in student composition.

## 4.2 Mechanisms

### 4.2.1 Environmental familiarity

Taking an exam in a non-home school exposes students to a setting that differs from their regular learning environment — new classrooms, layouts, peers, and proctors. These differences can impose additional cognitive burdens that make it harder for students to translate ability into performance (Smith, 1979; Nejati, 2023). In high-stakes settings such as the *Gaokao*, where effort and attention are already near their limits (Cai et al., 2019), even small disruptions in familiarity can noticeably affect performance.

**Table 2: Suggestive Evidence on Mechanisms**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Total Score | | | |
| Non-Home School | -0.148*** | -0.144*** | -0.126*** | -0.124*** |
| | (0.0190) | (0.0202) | (0.0182) | (0.0180) |
| × Any Nearby Classmate | 0.144* | | | |
| | (0.0821) | | | |
| × Number of Nearby Classmates (centered) | | 0.137** | | |
| | | (0.0710) | | |
| Travel Time | | | -0.00414*** | -0.00415*** |
| | | | (0.000607) | (0.000612) |
| Controls: | | | | |
| Any Nearby Classmate | X | . | . | . |
| Number of Nearby Classmates | . | X | . | . |
| Number of Adjacent Seats | X | X | . | . |
| Observations | 8,535 | 8,535 | 8,224 | 8,224 |
| R-squared | 0.513 | 0.513 | 0.493 | 0.501 |
| Individual Controls | . | . | . | X |
| Test Room FEs | X | X | X | X |
| Highschool Class FEs | X | X | X | X |

*Notes:* This table presents suggestive evidence on mechanisms. *Non-Home School* is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. *Any Nearby Classmate* is a dummy variable equal to 1 if a classmate is seated nearby. *Number of Nearby Classmates* measures the number of classmates seated nearby. *Number of Adjacent Seats* controls for the number of adjacent seats available to each examinee. *Travel Time* measures the travel time, in minutes, from the student's residential address to the test center by e-bike. *Total Score* represents the student's total test score, standardized by year and academic track. Standard errors are clustered at the high school class level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

To explore the familiarity mechanism, we first test whether social cues that reduce unfamiliarity — such as the presence of known peers — mitigate the performance penalty. As shown in Figure A3, each test room contains 30 seats arranged in five columns and six rows

[11] This pattern also suggests that the results are unlikely to be explained by the notion that non-home schools have poorer facilities that negatively impact performance.

under standardized exam protocols. To better capture the notion of familiarity, we construct an indicator for whether a student has a high-school classmate seated nearby (within one seat horizontally, vertically, or diagonally). Approximately 5 percent of students meet this criterion.[12] Because seats located on the edges or corners of the test room have fewer adjacent positions, we control for the number of adjacent seats available to each examinee to account for mechanical variation in seating geometry. Column (1) of Table 2 shows that having a nearby classmate significantly attenuates the performance penalty associated with non-home-school assessment. To further corroborate this result, Column (2) employs a continuous measure — the number of high-school classmates seated nearby — while again controlling for the total number of adjacent seats. Consistent with Column (1), the estimate indicates that an additional familiar peer substantially offsets the negative effect of non-home assessment.

Our second piece of suggestive evidence comes from examining subject-specific effects. If environmental unfamiliarity imposes additional cognitive burdens that disrupt concentration and task execution, its impact should be stronger in cognitively demanding subjects such as mathematics and science (Beilock and Carr, 2005). Table 3 presents the results. Consistent with this hypothesis, students assigned to a non-home center score 0.20 standard deviations lower in mathematics and 0.15 standard deviations lower in the comprehensive science test (which includes physics, chemistry, and biology). In contrast, the effects are smaller for language-based subjects — around 0.05 standard deviations in Chinese and English and 0.10 in the comprehensive liberal-arts test (which covers history, geography, and political science).[13] Although geography is formally classified as a liberal-arts subject, it includes some physical geography components that involve quantitative reasoning. This may explain the slightly larger penalty observed in the comprehensive liberal-arts test. Finally, Columns (6) – (7) in Table 3 show that the non-home penalty is more pronounced for students in the science track than for those in the liberal-arts track, though the difference is not statistically significant.[14] While more suggestive, these combined patterns are consistent with the environmental (un)familiarity explanation.

[12]We verify that whether a student has a classmate sitting nearby is uncorrelated with whether they are testing at their home school (due to the random assignment of seats), indicating that students at both home and non-home test centers have an equal likelihood of having a nearby classmate.

[13]The comprehensive tests have smaller samples because the science version is taken only by students in the science track, while the liberal-arts version is taken only by students in the liberal-arts track.

[14]Given that the performance penalty is concentrated in STEM subjects, Appendix Appendix B explores whether assessment location affects the' subsequent college or major choice of students.

**Table 3: Results by Subjects**

| | (1) Math | (2) Chinese | (3) English | (4) Comp. test (Science) | (5) Comp. test (Liberal Arts) | (6) Total Score | (7) Total Score |
|---|---|---|---|---|---|---|---|
| Non-Home School | -0.205*** | -0.0527** | -0.0486* | -0.153*** | -0.0977* | -0.151*** | -0.114* |
| | (0.0250) | (0.0238) | (0.0255) | (0.0205) | (0.0544) | (0.0198) | (0.0605) |
| Observations | 8,535 | 8,535 | 8,535 | 6,965 | 1,499 | 6,965 | 1,499 |
| R-squared | 0.333 | 0.325 | 0.325 | 0.504 | 0.422 | 0.578 | 0.411 |
| Test Room FEs | X | X | X | X | X | X | X |
| Highschool Class FEs | X | X | X | X | X | X | X |
| Academic Track | All | All | All | Science | Liberal Arts | Science | Liberal Arts |

*Notes:* This table presents our results by subjects. $Non\text{-}Home\ School$ is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. $Total\ Score$ represents the student's total test score, standardized by year and academic track. $Chinese$ represents the student's standardized test score on Chinese. $Math$ represents the student's standardized test score on Math. $English$ represents the student's standardized test score on English. $Comp.\ test\ (Science)$ refers to the student's standardized score on the comprehensive science exam, which includes physics, chemistry, and biology. $Comp.\ test\ (Liberal\ Arts)$ refers to the standardized score on the comprehensive liberal arts exam, covering history, geography, and political science. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

### 4.2.2 Travel distance

A second potential channel is travel distance to the assigned test center. Longer commutes may impose time and energy costs that reduce cognitive performance. Students who travel farther are more likely to experience shorter sleep durations, fatigue, or uncertainty about arriving on time — all of which can diminish focus during the exam. In our context, the *Gaokao* begins at 9 a.m., and students must arrive 45–60 minutes early for identity verification and security checks. Anecdotally, many students find it difficult to sleep the night before the exam, making any additional commute time that cuts into rest particularly consequential. Consistent with this reasoning, Heissel and Norris (2018) show that school start times and sleep duration have sizable effects on academic performance.

To investigate this mechanism, we calculate the travel time between each student's home and their assigned test center using e-bike transportation, the most common mode of transportation in our sampled county.[15] On average, it takes a student 24.6 minutes to travel from home to the test center, with the 25th percentile at 13 minutes and the 75th percentile at 35 minutes. We include travel time in Column (3) of Table 2 and make two observations: (1)

[15]The calculation is performed by a local government agent using our code, which relies on the AMap API — the Chinese equivalent of Google Maps. Due to changes in street names, however, travel time information is unavailable for around 300 students.

longer travel time is associated with poorer test performance; (2) including travel time does not meaningfully alter the magnitude of the non-home school effect. These results imply that travel-related costs contribute to but do not largely explain the performance penalty.

The modest role of travel time likely reflects the fact that test centers in our county are geographically concentrated. As shown in Figure A2, most students live within the central urban area, and those assigned to non-home schools travel only slightly farther than students who remain at their home school (see Figure A4). Column (4) of Table 2 adds controls for socioeconomic status, proxied by housing prices, with similar results.[16]

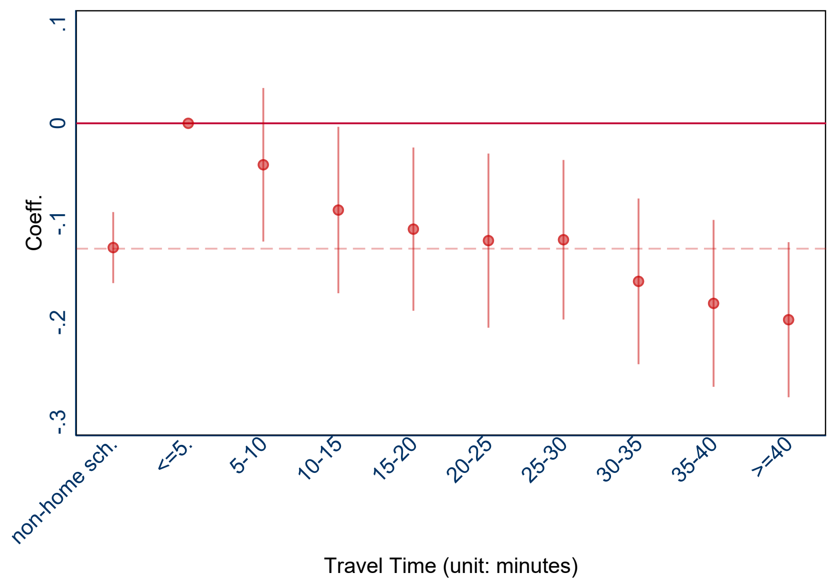
To further illustrate the relationship between travel time and performance, we estimate a flexible specification that replaces continuous travel time with 5-minute interval dummies, using students within 5 minutes of their test center as the reference group. We also include our main variable of interest — non-home exam-taking — in this specification. Figure 2 shows a clear monotonic decline in scores with increasing travel time, with the largest penalty among students traveling the farthest. Quantitatively, the effect of testing in a non-home school is roughly equivalent to the performance loss associated with an additional 25 minutes of travel.

#### 4.2.3 Cheating and other factors

A natural concern is that taking the exam at one's home school may facilitate cheating. However, cheating is unlikely to drive our results given the strict anti-cheating protocols in place. During our sample period, the NCEE operated under real-time monitoring systems and rigorous proctoring procedures, making large-scale cheating extremely difficult (Borcan, Lindahl and Mitrut, 2017). Each test room is supervised by two proctors recruited from outside the county, and proctors are reassigned for each subject, with assignments determined randomly immediately before the exam. Monitoring cameras are installed in every test room and recorded the entire exam process in real time under the supervision of provincial authorities. Figure A5 provides an illustration. To further enhance exam integrity, additional protocols have been implemented during the NCEE to prevent cheating: (1) while students are permitted to use the restroom during the exam, few do so due to time constraints, and those who do are accompanied and monitored by a same-sex proctor; (2) students are prohibited from wearing school uniforms to prevent identification with specific high schools; (3) students are prohibited from bringing any electronic devices into the test room; (4) all examinees are in-

[16]Students who live far from test centers may mitigate travel costs by staying in nearby accommodations. In the absence of such data, our estimates of travel-time effects are likely conservative.

**Figure 2: Travel Time and Test Performance**




*Notes:* This figure illustrates the relationship between travel time and test performance, using students who live very close to the test center (travel time less than 5 minutes) as the baseline group. 95% confidence intervals based on high school class clusters are reported.

dividually screened with metal detectors upon entry, and radio signal jammers are used to ensure full and effective coverage of the test rooms; (5) all desks and chairs are inspected before the start of each exam session; and (6) nearby roads are closed to motor vehicles in order to reduce the noise level. Since 2015, organizers convicted of facilitating mass cheating face up to seven years in prison, and no such cases were reported in our samples.

Quantitatively, our results are consistently significant across test centers and years. Even if some isolated cases of cheating existed, they are unlikely to be the main mechanism behind our findings. Moreover, it is difficult to reconcile the stronger mitigating effect of having a nearby classmate in non-home test centers with a cheating-based explanation — conceptually, familiar peers in home test centers would be more conducive to cheating.

Another potential concern is that certain room or seating positions might offer performance advantages, and that students taking the exam at their home school could leverage prior familiarity to secure "better" seats (Li and Zhou, 2026). However, test room and seating assignments are also randomly determined before the exam and fixed for the entire testing period, leaving no scope for strategic selection advantages.

Finally, we discuss whether other short-term environmental factors could explain our findings. A large body of research shows that temporary environmental conditions — such as temperature, air pollution, and pollen — can influence test performance (Marcotte, 2015; Ebenstein, Lavy and Roth, 2016; Bensnes, 2016; Graff Zivin, Hsiang and Neidell, 2018; Graff Zivin et al., 2020; Park, 2022). However, these factors are unlikely to account for our results. First, all test centers in our sample are located in the central part of the county, a relatively compact area with limited variation in environmental conditions. Second, our preferred specification includes test-room fixed effects, which absorb environmental differences at a highly granular level, including room-specific characteristics such as floor level, facility quality, and other location-specific attributes.

Overall, the evidence suggests that environmental unfamiliarity may likely be a main mechanism of the off-site assessment penalty, while travel distance plays a secondary role. Logically, taking the exam in an unfamiliar environment imposes additional cognitive burdens that temporarily make it harder for students to demonstrate their full ability in a high-stakes setting. These findings indicate that assessment performance reflects not only individuals' underlying ability, but also the conditions under which that ability is realized.

## 4.3 Further results on disparity

In this section, we document two dimensions of inequality to better understand its implications: disadvantaged students not only experience larger performance penalties when exposed to unfavorable testing environments but are also more likely to be assigned to them, reflecting systematic differences in how test centers are designated.

**Heterogeneity by Student Background**. We begin by examining whether low-performing students are more sensitive to unfamiliar environments than their high-performing peers. Students with higher academic ability may possess greater self-efficacy and task focus, making them less susceptible to unfamiliar testing environments. Students are classified based on their total exam scores, with those scoring above the median defined as high achievers.[17] The results are presented in Columns (1) and (2) of Panel A, Table 4. We find that our main results are

[17] Our classification may be endogenous to test center assignments, as it is based on realized test scores from the college entrance exam. However, this approach is likely to introduce only moderate measurement error. As shown in Table 1, taking the exam at a non-home institution alters a student's rank percentile by an average of just 3 points. To ensure robustness, we also adopt an alternative definition of high and low achievers by excluding students ranked in the 45th to 55th percentiles. Specifically, we define high achievers as those ranked in the 0–45th percentile and low achievers as those in the 55th–100th percentile. The results remain similar under this alternative classification.

primarily driven by low achievers: their total test scores are 0.175 standard deviations lower when they take the exam in a non-home school. In contrast, the negative effect of a non-home school setting is substantially smaller for high achievers.

**Table 4: Heterogeneity Analysis**

| | Panel A | | | | | |
|---|---|---|---|---|---|---|
| | Total Score | | Total Score | | Exam Retake | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Sample | Low achievers | High achievers | Low SES | High SES | Low SES | High SES |
| Non-Home School | -0.175*** | -0.0313** | -0.171*** | -0.101*** | 0.0281 | 0.237*** |
| | (0.0269) | (0.0152) | (0.0297) | (0.0270) | (0.0171) | (0.0213) |
| Statistical difference | P-value<0.001*** | | P-value=0.076* | | P-value<0.001*** | |
| Observations | 4,135 | 4,379 | 4,070 | 4,459 | 2,733 | 2,882 |
| R-squared | 0.368 | 0.450 | 0.569 | 0.555 | 0.169 | 0.198 |
| Test Room FEs | X | X | X | X | X | X |
| Highschool Class FEs | X | X | X | X | X | X |
| Mean of dep. var. | -0.569 | 0.890 | 0.128 | 0.232 | 0.110 | 0.280 |

| | Panel B | | | | | |
|---|---|---|---|---|---|---|
| | College Admission | | College Admission | | Final College Admission | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Sample | Low achievers | High achievers | Low SES | High SES | Low SES | High SES |
| Non-Home School | -0.0453** | -0.00363 | -0.0428*** | -0.0328** | -0.0317* | 0.0118 |
| | (0.0201) | (0.00763) | (0.0146) | (0.0155) | (0.0194) | (0.0198) |
| Statistical difference | P-value=0.034** | | P-value=0.240 | | P-value=0.074* | |
| Observations | 4,135 | 4,379 | 4,070 | 4,459 | 2,733 | 2,882 |
| R-squared | 0.461 | 0.479 | 0.544 | 0.497 | 0.493 | 0.407 |
| Test Room FEs | X | X | X | X | X | X |
| Highschool Class FEs | X | X | X | X | X | X |
| Mean of dep. var. | 0.384 | 0.947 | 0.658 | 0.689 | 0.685 | 0.765 |

*Notes:* This table presents our additional results. $Non\text{-}Home\ School$ is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. Student socioeconomic status (SES) is classified based on whether the average housing price in their neighborhood falls above/below the county median of ¥2,500 per square meter. $Total\ Score$ represents the student's total test score, standardized by year and academic track. $Exam\ Retake$ is a binary variable indicating whether student $i$ retook the exam in the following year. $College\ Admission$ is a binary variable indicating whether student $i$ is admitted to any college in the current year. $Final\ College\ Admission$ is a binary variable indicating whether student $i$ is admitted to any college either in the current year or the following year if they retook the exam. Standard errors are clustered at the high school class level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Furthermore, we investigate whether the performance penalty associated with the exam location differs for students from different socioeconomic backgrounds. Although students do not report their household income when registering for the exam, they are required to provide their home address in order to receive a potential offer letter. We match their home address to the average price of the neighborhood and classify students as belonging to a low socioeconomic class if the price of their neighborhood is below the median house price (i.e., ¥2500 per square meter). The results, presented in Columns (3) – (4) of Panel A in Table 4, show that students from low SES are more strongly affected by testing in a non-home school.

We then examine the heterogeneity in exam retaking (Goodman, Gurantz and Smith, 2020). While Table 1 shows that students assigned to non-home schools are more likely to be dissatisfied with their scores and retake the exam the following year, this average effect masks substantial heterogeneity across socioeconomic groups. Retaking the NCEE involves substantial costs, both in terms of time and financial resources.[18] Columns (5) – (6) of Panel A in Table 4 shows that the retaking is largely driven by students from higher socioeconomic backgrounds, who are 23.7 percentage points more likely to retake the exam when assigned to a non-home test center. In contrast, students from lower socioeconomic backgrounds are only 2.8 percentage points more likely to retake the exam, likely due to financial constraints imposed by the high costs associated with retaking.[19]

We close here by exploring the implications of inequality on college admission outcomes. Panel B of Table 4 presents the results. In Column (1), we find that low achievers are 4.53 percentage points less likely to be admitted to college if assigned to a non-home test center. However, high-achieving students are not affected in terms of college admission, likely because they are not marginal students near the admission cutoff and their performance is less sensitive to assessment location. Moreover, in Columns (3) and (4), we find that students from low SES are more adversely affected by testing in a non-home school with respect to college admission. In Columns (5) and (6), we further account for exam retakes by constructing a new variable — final college admission — which captures admission outcomes both in the current year and the following year for exam retakers. We find that when accounting for retakes, high-SES students are much less affected by their initial test center assignment, likely because they have more resources to retake the exam if unsatisfied with their first attempt. However, for low-SES students, the initial test center assignment still significantly influences their final college

[18] Since the exam is administered only once a year, students must spend an additional year preparing. Public high schools do not admit NCEE retakers, so these students typically enroll in private schools, which cost ¥15,000 in our sample county — where GDP per capita was approximately ¥25,000 in 2016. In addition to tuition, retakers forgo income they could have earned during that year, a burden that may be particularly significant for students from low SES.

[19] Unlike SAT-style tests — where students can submit their highest score, and retaking generally increases admission chances — the NCEE system is more uncertain. Retaking the NCEE means competing against a new cohort of students with new scores, and even with improved performance, better college admission outcomes are not guaranteed.

admission outcomes.

**Uneven Distribution of Test Centers.** Similar to many other contexts, test centers in China are concentrated in more advantaged high schools. Table A5 presents a quantitative comparison of high schools that serve as test centers and those that do not. Students from high schools that do not serve as test centers are more likely to come from disadvantaged backgrounds: they are more likely to come from low-SES families, must travel greater distances to reach test centers, exhibit significantly poorer academic performance, and are less likely to be admitted to any (elite) college.

It is important to note that our main analysis focuses only on students from high schools designated as test centers. While this approach provides meaningful variation for the purpose of causal identification, it excludes students from high schools that do not serve as test centers, who may be even more vulnerable to testing in an unfamiliar environment. In fact, the average performance of students from non-test-center high schools is slightly lower than that of low-achieving students — those below the median — from test-center high schools (-0.59 vs. -0.57). Therefore, our baseline estimates likely represent a lower bound of the overall performance penalty associated with non-home test locations.

Lastly, we conduct a simple back-of-the-envelope calculation to assess the extent to which the performance gap between students from test-center high schools and those from non-test-center schools can be explained by exam location. As shown in Table A5, the performance gap is 0.767 (0.182 v.s. -0.586). In our sample county, 82.9% of students from test-center high schools took the exam at a non-home school and on average experience a performance decline by 0.140. On the other hand, 100% of students from non–test-center high schools took the exam at a non-home school. Because this group lacks meaningful variation in test location, we cannot directly estimate the effect of testing away from home for them. Instead, we assume that the penalty for students from non-test-center schools is similar to that of low-achieving students in test-center schools as an approximate estimate — a performance decline of 0.175 (Table 4).[20] In a counterfactual scenario where all students take the exam at their home school, those from test-center high schools would score an average of 0.298 (computed as $0.182+0.140\times82.9\%$), while students from non-test-center high schools would score $-0.411$ (computed as $-0.586+0.175\times100\%$), resulting in a performance gap of 0.709. Therefore,

[20] We acknowledge that our current approximate approach may not fully capture the exact penalty, and that the true magnitude could be larger, as students from non-test-center high schools also tend to travel longer distances and are more likely to come from lower-SES families.

7.6% of the observed performance gap could be attributed to the exam location (computed as $\frac{0.767-0.709}{0.767}$).

## 4.4 Global relevance and implications

While our identification relies on the random assignment of test centers in China's NCEE, the mechanisms we uncover — environmental unfamiliarity and travel-induced cognitive load — appear across many high-stakes assessment systems worldwide. These forces generally reduce measured performance, although the magnitude of the effect can naturally vary with institutional design, exam stakes, and local logistical conditions.

Different institutional arrangements manifest these frictions in distinctive ways. In centralized systems that prioritize procedural fairness and anonymity, such as South Korea and France, students are intentionally assigned to external venues. In South Korea, for example, candidates learn their designated test centers only one day before the College Scholastic Ability Test, and they are not permitted to enter the assigned classrooms in advance. This leaves them no opportunity to familiarize themselves with the testing environment. Such practices illustrate an implicit institutional trade-off in which test security is achieved at the cost of students' cognitive comfort. Although this "forced unfamiliarity" is widely regarded as standard administrative procedure, our findings provide causal evidence of the performance penalty embedded in this displacement. The results suggest that rules designed to appear neutral may systematically disadvantage students who are less able to adapt to new environments. In large-scale examinations in emerging economies such as Brazil (ENEM) and India (JEE/NEET), the challenges associated with off-site assessment are intensified by logistical constraints. Centralized allocation in these settings often requires candidates, particularly those from rural or underserved areas, to travel substantial distances to urban test centers. This adds physical fatigue to the cognitive demands posed by an unfamiliar venue.

In decentralized or market-driven systems, the barrier is often structural. The scarcity of assessment centers in low-income neighborhoods in the United States creates "testing deserts" that necessitate travel. On a global scale, high-stakes professional and graduate assessments (such as the GRE or CFA) rely on commercial testing networks concentrated in major metropolitan hubs, effectively imposing a "rural tax" on candidates from non-urban regions. In the case of Japan, the barrier is even more stringent: university-specific entrance examinations require students to navigate to particular college campuses, conferring an inherent advantage upon local students familiar with the environment.

Viewing assessment location as part of the measurement infrastructure rather than as a neutral administrative detail opens new avenues for policy. Reducing travel burdens, expanding access to familiar testing environments, and increasing school-based test administration may improve both the accuracy and the equity of high-stakes assessments. Recognizing this, some policymakers have, in practice, taken steps to improve equal access through better test design. Since 2000, several U.S. states have implemented mandates requiring high school juniors to take a college entrance exam (e.g., the SAT or ACT), administered during school hours at students' home schools (Klasik, 2013; Hurwitz et al., 2015; Goodman, 2016; Hyman, 2017). In Brazil, the education authority has recently refined its assignment algorithm to shorten the distance between students' homes and their test centers.[21]

More broadly, our results underscore a conceptual point: improving human capital is only part of the policy challenge; societies must also be able to measure it accurately. When assessments take place in systematically unequal or unfamiliar environments, measured performance may conflate true ability with contextual conditions. This affects both individual opportunity and the broader informational role that assessments play in human-capital development, especially for disadvantaged groups.

# 5 Concluding Remarks

The context in which ability is measured matters. Despite growing policy debates, we lack causal evidence on the role of assessment location ("where") in shaping high-stakes cognitive performance. In this paper, we address this question by leveraging China's college entrance exam — a life-determining high-stakes assessment — which randomly assigns students to different test centers. Using administrative data in a large county, we find that students assigned to a non-home school score 0.14 standard deviations lower than those taking the exam at their home school, thereby reducing their likelihood of college admission. We show suggestive evidence that the observed effect is likely to be primarily driven by environmental unfamiliarity that can impose additional cognitive burdens: (i) the presence of a familiar peer nearby mitigates the effect; (ii) the effect is most pronounced in cognitively demanding STEM subjects; and (iii) controlling for travel distance does not substantially alter the estimate. Moreover, the impact is disproportionately large among low-performing and low-SES individuals, which may further exacerbate inequality in access to future opportunities. Crucially, we document

[21] See https://www.gov.br/inep/pt-br/centrais-de-conteudo/noticias/enem/inep-estuda-novas-medidas-para-aperfeicoar-logistica-do-exame.

that these initial measurement distortions translate into real economic scarring: affected students experience a discernible penalty in early-career earnings and exhibit reduced geographic mobility in the labor market.

Collectively, our findings speak to broader questions about accurately assessing human capital in high-stakes contexts and the opportunities allocated as a result. To mitigate performance biases associated with assessment location, policymakers may consider improving disadvantaged groups' accessibility to test centers where administratively feasible. Since our analysis suggests that environmental familiarity may be an important mechanism, providing examinees with greater opportunities to become familiar with their testing environment may also help. Finally, emerging technological solutions — such as the increasing feasibility of proctored at-home or online testing — may offer promising avenues for reducing these contextual barriers. Future research could examine whether such innovations can lessen cognitive disruption in high-stakes assessments and assess their potential to narrow inequality.

# References

**Anelli, Massimo.** 2020. "The Returns to Elite University Education: a Quasi-Experimental Analysis." *Journal of the European Economic Association*, 18(6): 2824–2868.

**Angrist, Joshua D., and Victor Lavy.** 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement*." *The Quarterly Journal of Economics*, 114(2): 533–575.

**Beilock, Sian L, and Thomas H Carr.** 2005. "When high-powered people fail: Working memory and "choking under pressure" in math." *Psychological science*, 16(2): 101–105.

**Bensnes, Simon Søbstad.** 2016. "You sneeze, you lose:: The impact of pollen exposure on cognitive performance during high-stakes high school exams." *Journal of Health Economics*, 49: 1–13.

**Biasi, Barbara, Julien Lafortune, and David Schönholzer.** 2025. "What Works and for Whom? Effectiveness and Efficiency of School Capital Investments Across the U.S.*." *The Quarterly Journal of Economics*, 140(3): 2329–2379.

**Bond, Timothy N., Jillian B. Carr, Analisa Packham, and Jonathan Smith.** 2022. "Hungry for Success? SNAP Timing, High-Stakes Exam Performance, and College Attendance." *American Economic Journal: Economic Policy*, 14(4): 51–79.

**Borcan, Oana, Mikael Lindahl, and Andreea Mitrut.** 2017. "Fighting Corruption in Education: What Works and Who Benefits?" *American Economic Journal: Economic Policy*, 9(1): 180–209.

**Bulman, George.** 2015. "The effect of access to college assessments on enrollment and attainment." *American Economic Journal: Applied Economics*, 7(4): 1–36.

**Cai, Xiqian, Yi Lu, Jessica Pan, and Songfa Zhong.** 2019. "Gender Gap under Pressure: Evidence from China's National College Entrance Examination." *The Review of Economics and Statistics*, 101(2): 249–263.

**Chang, Eunsik, and Maria Padilla-Romo.** 2023. "When Crime Comes to the Neighborhood: Short-Term Shocks to Student Cognition and Secondary Consequences." *Journal of Labor Economics.*

**Chetty, Raj, John N. Friedman, and Jonah E. Rockoff.** 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–79.

**Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz.** 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106(4): 855–902.

**Chiang, Jeannine.** 2024. "Why Are Some Students Traveling Hundreds of Miles to Take the SAT?" *The Nation.*

**Ebenstein, Avraham, Victor Lavy, and Sefi Roth.** 2016. "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution." *American Economic Journal: Applied Economics*, 8(4): 36–65.

**Franco, Catalina, and Erika Povea.** 2025. "Exam Luck and Human Capital Accumulation." *Working Paper.*

**Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek.** 2012. "Long-Term Effects of Class Size." *The Quarterly Journal of Economics*, 128(1): 249–285.

**Gaggero, Alessio, and Denni Tommasi.** 2022. "Time of Day and High-Stake Cognitive Assessments." *The Economic Journal*, 133(652): 1407–1429.

**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.

**Goodman, Sarena.** 2016. "Learning from the Test: Raising Selective College Enrollment by Providing Information." *The Review of Economics and Statistics*, 98(4): 671–684.

**Graff Zivin, Joshua, Solomon M. Hsiang, and Matthew Neidell.** 2018. "Temperature and Human Capital in the Short and Long Run." *Journal of the Association of Environmental and Resource Economists*, 5(1): 77–105.

**Graff Zivin, Joshua, Tong Liu, Yingquan Song, Qu Tang, and Peng Zhang.** 2020. "The unintended impacts of agricultural fires: Human capital in China." *Journal of Development Economics*, 147: 102560.

**Heissel, Jennifer A, and Samuel Norris.** 2018. "Rise and shine: The effect of school start times on academic performance from childhood through puberty." *Journal of Human Resources*, 53(4): 957–992.

**Hoekstra, Mark.** 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *The Review of Economics and Statistics*, 91(4): 717–724.

**Hoxby, Caroline M.** 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation*." *The Quarterly Journal of Economics*, 115(4): 1239–1285.

**Huang, Hai, Wei Huang, Xinzheng Shi, and Ming ang Zhang.** 2025. "Gender composition, social context, and academic performance in high-stakes examinations." *Journal of Development Economics*, 176: 103509.

**Hurwitz, Michael, Jonathan Smith, Sunny Niu, and Jessica Howell.** 2015. "The Maine Question: How Is 4-Year College Enrollment Affected by Mandatory College Entrance Exams?" *Educational Evaluation and Policy Analysis*, 37(1): 138–159.

**Hyman, Joshua.** 2017. "ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice." *Education Finance and Policy*, 12(3): 281–311.

**Jackson, C. Kirabo, and Claire L. Mackevicius.** 2024. "What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States." *American Economic Journal: Applied Economics*, 16(1): 412–46.

**Jia, Ruixue, and Hongbin Li.** 2021. "Just above the exam cutoff score: Elite college admission and wages in China." *Journal of Public Economics*, 196: 104371.

**Jia, Ruixue, Hongbin Li, and Claire Cousineau.** 2025. *The Highest Exam: How the Gaokao Shapes China.* Harvard University Press.

**Kang, Le, Ziteng Lei, Yang Song, and Peng Zhang.** 2024. "Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes." *Journal of Political Economy: Microeconomics*, 2(2): 355–397.

**Klasik, Daniel.** 2013. "The ACT of Enrollment: The College Enrollment Effects of State-Required College Entrance Exam Testing." *Educational Researcher*, 42(3): 151–160.

**Krueger, Alan B.** 1999. "Experimental Estimates of Education Production Functions*." *The Quarterly Journal of Economics*, 114(2): 497–532.

**Landaud, Fanny, Éric Maurin, Barton Willage, and Alexander Willén.** 2024. "The value of a high school GPA." *Review of Economics and Statistics*, 1–24.

**Lee, Steven, and Matthew Schaelling.** 2024. "Content Relatability and Test Score Disparities: Evidence from Texas." *Working Paper.*

**Li, Hongbin, and Xinyao Qiu.** 2023. "Heuristics in Self-Evaluation: Evidence from the Centralized College Admission System in China." *The Review of Economics and Statistics*, 1–36.

**Li, Hongbin, Lingsheng Meng, Kai Mu, and Shaoda Wang.** 2024. "English language requirement and educational inequality: Evidence from 16 million college applicants in China." *Journal of Development Economics*, 168: 103271.

**Li, Xuan, and Xiang Zhou.** 2026. "The cognitive benefit of a window view." *Journal of Economic Behavior & Organization*, 241: 107358.

**Marcotte, Dave E.** 2015. "Allergy test: Seasonal allergens and performance in school." *Journal of health economics*, 40: 132–140.

**Nejati, Vahid.** 2023. "Chapter two - Primary principles of cognitive rehabilitation." In *Principles of Cognitive Rehabilitation.* , ed. Vahid Nejati, 59–137. Academic Press.

**Park, R Jisung.** 2022. "Hot Temperature and High-Stakes Performance." *Journal of Human Resources*, 57(2): 400–434.

**Sekhri, Sheetal.** 2020. "Prestige Matters: Wage Premium and Value Addition in Elite Colleges." *American Economic Journal: Applied Economics*, 12(3): 207–25.

**Smith, Kestrel A.** 2018. *Invisible Barriers to Higher Education in Indian Country: Standardized Testing*. The University of Arizona.

**Smith, Steven M.** 1979. "Remembering in and out of context." *Journal of Experimental Psychology: Human Learning and Memory*, 5(5): 460.

**Wang, Ao, Shaoda Wang, and Xiaoyang Ye.** 2023. "When Information Conflicts with Obligations: the Role of Motivated Cognition." *The Economic Journal*, 133(654): 2533–2552.

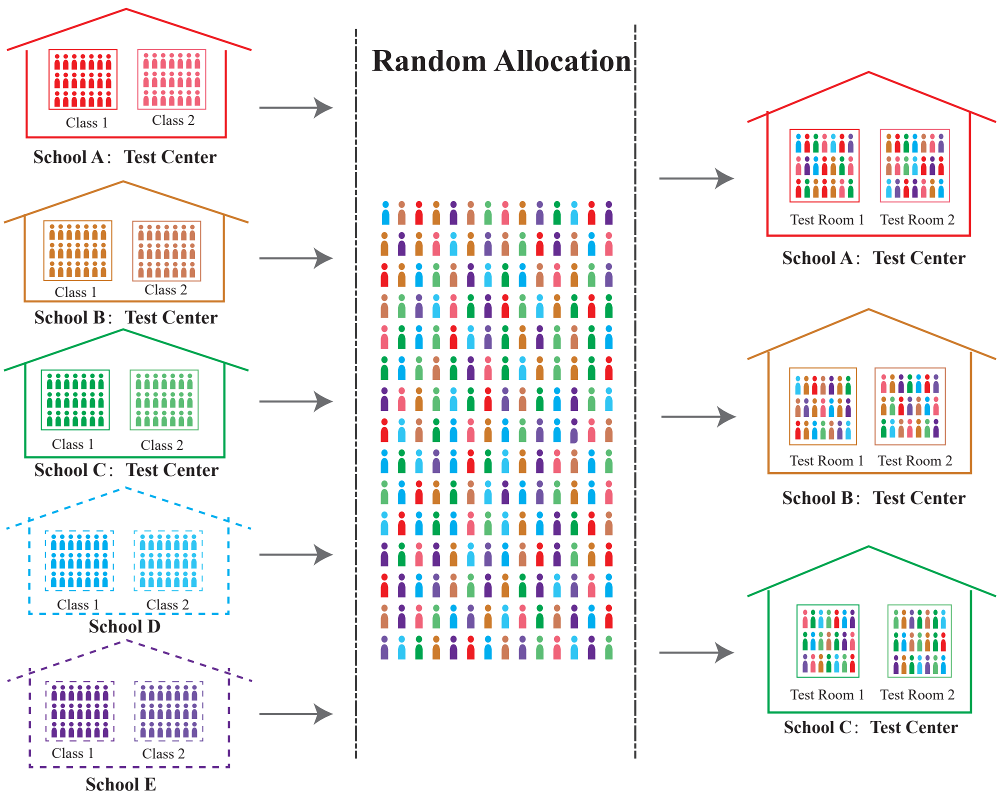
**Zimmerman, Seth D.** 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics*, 32(4): 711–754.

**Zimmerman, Seth D.** 2019. "Elite Colleges and Upward Mobility to Top Jobs and Top Incomes." *American Economic Review*, 109(1): 1–47.

# Appendices

## Appendix A Additional Figures and Tables

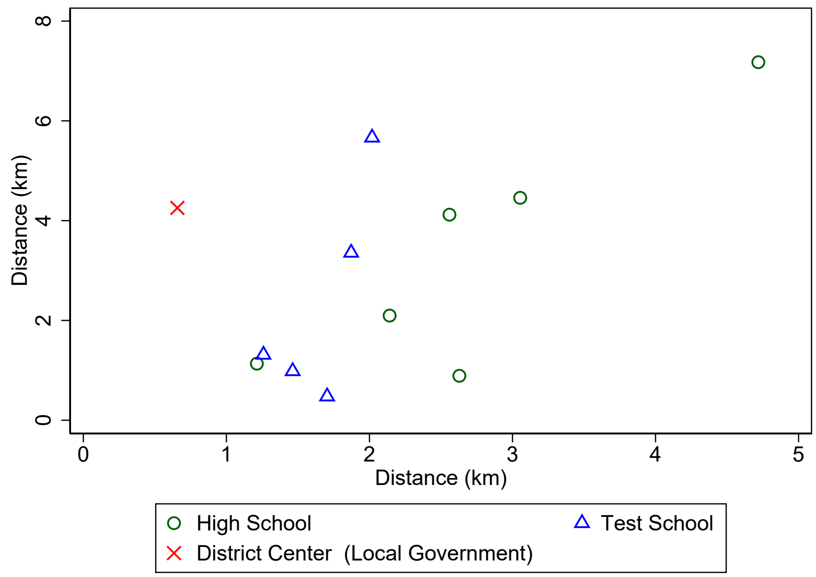
**Figure A1: Random Allocation of Students to Test Centers**




*Notes:* This figure illustrates the process of randomly allocating students to test centers. First, all students within the same county — whether from high schools that host test centers or those that do not — are pooled together. They are then randomly assigned across the available test centers, meaning that students from the same high school class may end up taking the exam in different locations.

[Back to Page 3]

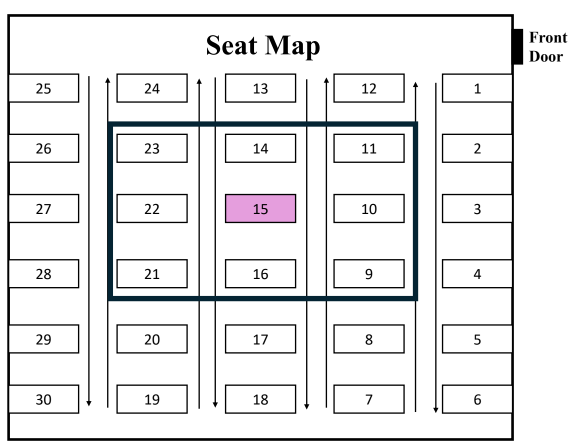
**Figure A2: Relative Locations of High Schools in the Sample County**




*Notes:* This figure depicts the *relative* locations of high schools and designated test centers within the sample county, since disclosure of the actual map is not permitted. Green circles represent high schools that do not serve as test centers, while blue triangles indicate the five that serve as test centers. The red cross marks the location of the local government. The origin point is simply chosen for ease of visualization. Both axes measure distance in kilometers, illustrating the relative positions of these locations.

[Back to Page 9]

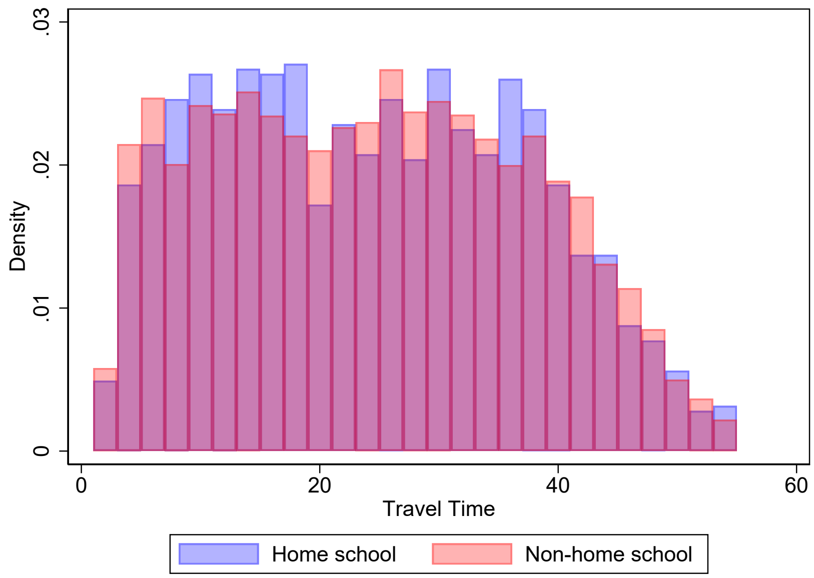
**Figure A3: Test Room Floor Plan**




*Notes:* The figure shows the seating arrangement of a standardized test room used in the study. The room is organized into five columns, each consisting of six rows of desks. For example, for seat No. 15, we define seats 9, 10, 11, 14, 16, 21, 22, and 23 as its nearby seats.

[Back to Page 18]

**Figure A4: Distribution of Travel Time**




*Notes:* The figure shows travel-time distributions for students assigned to home *versus* non-home schools. The sample is restricted to students from the five high schools that serve as test centers, consistent with the sample used in our regression analysis. The x-axis represents travel time, in minutes, from the student's residential address to the test center by e-bike.

[Back to Page 21]

**Figure A5: Real-Time Monitoring System**


*Notes:* The figure shows the real-time monitoring system used in NCEE. Source: *China Daily*.

[Back to Page 21]

**Table A1: Socioeconomic Characteristics of the Sampled County**

| | GDP (10,000 CNY) | Fiscal income (10,000 CNY) | Fiscal expenditure (10,000 CNY) | High schoolers (share of pop.) | Schooler-Teacher ratio in high schools |
|---|---|---|---|---|---|
| **Year: 2016** | | | | | |
| Sampled county | 2,753,981 | 81,094 | 481,908 | 4.3% | 14.5 |
| Average same-province county | 2,760,522 | 110,193 | 351,205 | 4.4% | 15.4 |
| Average Chinese county | 2,235,025 | 167,878 | 344,827 | 4.4% | 13.7 |
| **Year: 2017** | | | | | |
| Sampled county | 2,820,765 | 85,413 | 509,171 | 4.3% | 14.5 |
| Average same-province county | 2,922,147 | 109,135 | 377,517 | 4.5% | 15.1 |
| Average Chinese county | 2,377,611 | 175,838 | 375,689 | 4.5% | 13.4 |
| **Year: 2018** | | | | | |
| Sampled county | 2,913,472 | 90,765 | 557,982 | 4.5% | 14.3 |
| Average same-province county | 2,922,147 | 109,135 | 377,517 | 4.5% | 14.7 |
| Average Chinese county | 2,583,394 | 184,745 | 409,540 | 4.6% | 13.1 |

*Notes:* This table presents the socioeconomic conditions of our sampled county. For comparison, we also provide corresponding statistics for an average county in the same province and for an average county in China. Data source: national, provincial, and county statistical yearbooks.

[Back to Page 9]

**Table A2: Balance Checks**

| | (1) Male (binary) | (2) Student age | (3) Urban household (binary) | (4) Class monitor (binary) | (5) Science track (binary) | (6) High house price (binary) | (7) Any classmate in the same test room (binary) |
|---|---|---|---|---|---|---|---|
| Non-Home School | 0.0146 | -0.00345 | -0.00582 | -2.90e-05 | 0.00602 | -0.0219 | -0.00586 |
| | (0.0144) | (0.0124) | (0.0140) | (0.00713) | (0.0112) | (0.0143) | (0.0118) |
| Observations | 8,535 | 8,535 | 8,535 | 8,535 | 8,535 | 8,535 | 8,535 |
| R-squared | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Mean of dep. var. | 0.493 | 17.59 | 0.386 | 0.0657 | 0.817 | 0.523 | 0.211 |

*Notes:* This table presents balance checks for various demographic and background variables, comparing students taking exams in home schools versus non-home schools. $Non\text{-}Home\ School$ is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. Robust standard errors are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[Back to Page 13]

**Table A3: Additional Results**

| | Panel A: Additional Outcomes | | |
|---|---|---|---|
| | (1) Raw Total Score | (2) Technically-Oriented College | (3) Stem Major |
| Non-Home School | -10.13*** | 0.0105 | -0.0161 |
| | (1.390) | (0.0139) | (0.0168) |
| Observations | 8,535 | 7,848 | 5,211 |
| R-squared | 0.554 | 0.182 | 0.139 |
| Test Room FEs | X | X | X |
| Highschool Class FEs | X | X | X |
| Cluster level | Highschool Class | Highschool Class | Highschool Class |
| Sample | Main Sample | College-Admitted Students | College-Admitted Students from the Science Track |

| | Panel B: Additional Specifications | | |
|---|---|---|---|
| | (1) Total Score | (2) Total Score | (3) Total Score |
| Non-Home School | -0.141*** | -0.141*** | -0.153*** |
| | (0.0214) | (0.0192) | (0.00440) |
| Total Score (Last Year) | | | 0.975*** |
| | | | (0.00902) |
| Observations | 8,535 | 8,535 | 1,065 |
| R-squared | 0.512 | 0.512 | 0.996 |
| Test Room FEs | X | X | . |
| Highschool Class FEs | X | X | . |
| School-Track-Year FEs | . | . | X |
| Cluster level | High school × Year × Track | Test Room | High school × Year × Track |
| Sample | Main Sample | Main Sample | Repeat Exam Takers |

*Notes:* This table presents our additional results. $Non\text{-}Home\ School$ is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. In Panel A, Column (1) uses raw total test scores as the outcome variable, while Columns (2) and (3) use college admission to technically-oriented programs and whether the major is in STEM, respectively. Standard errors are clustered at the high school class level. In Panel B, $Total\ Score$ represents the student's total test score, standardized by year and academic track. $TotalScore(LastYear)$ refers to an individual's standardized total test score from the previous year. Standard errors are clustered at the school-year-track level in Columns (1) and (3), and at the test room level in Column (2). *** $p$<0.01, ** $p$<0.05, * $p$<0.1.

[Back to Page 15] [Back to Page 42]

**Table A4: Additional Results by Subgroups**

| | Panel A: Results by Year and Gender | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | | Total Score | | Total Score | |
| Sample | 2016 | 2017 | 2018 | Male | Female |
| Non-Home School | -0.134*** | -0.125*** | -0.164*** | -0.152*** | -0.129*** |
| | (0.0368) | (0.0340) | (0.0260) | (0.0272) | (0.0263) |
| Observations | 2,762 | 2,857 | 2,916 | 4,203 | 4,328 |
| R-squared | 0.507 | 0.508 | 0.520 | 0.613 | 0.508 |
| Test Room FEs | X | X | X | X | X |
| Highschool Class FEs | X | X | X | X | X |

| | Panel B: Results by High School | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| | | | Total Score | | |
| Sample | School A | School B | School C | School D | School E |
| Non-Home School | -0.108*** | -0.127*** | -0.146*** | -0.107*** | -0.202*** |
| | (0.0310) | (0.0429) | (0.0459) | (0.0380) | (0.0395) |
| Observations | 2,283 | 1,746 | 1,661 | 1,208 | 1,637 |
| R-squared | 0.428 | 0.424 | 0.447 | 0.390 | 0.484 |
| Highschool Class FEs | X | X | X | X | X |
| Mean of "Non-Home School" | 0.823 | 0.848 | 0.827 | 0.801 | 0.839 |

*Notes:* This table presents our additional results by subgroups. $Non\text{-}Home\ School$ is a binary variable indicating whether student $i$ is assigned to a test center outside their home school to take the exam. $Total\ Score$ represents the student's total test score, standardized by year and academic track. Standard errors are clustered at the high school class level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[Back to Page 17]

**Table A5: Summary Statistics for High Schools**

| | High schools as test centers | High schools not as test centers | Mean difference |
|---|---|---|---|
| Panel A: Student characteristics | | | |
| Male | 0.493 | 0.454 | 0.039*** |
| | (0.500) | (0.498) | (0.000) |
| Han ethnicity | 0.997 | 0.998 | -0.001 |
| | (0.053) | (0.047) | (0.552) |
| Age | 17.588 | 17.594 | -0.006 |
| | (0.436) | (0.414) | (0.010) |
| Urban household | 0.386 | 0.324 | 0.061*** |
| | (0.487) | (0.468) | (0.000) |
| High-housing-price community | 0.523 | 0.485 | 0.038*** |
| | (0.500) | (0.500) | (0.000) |
| Science track | 0.817 | 0.501 | 0.316*** |
| | (0.387) | (0.500) | (0.000) |
| Distance to test centers | 6.144 | 6.308 | -0.163** |
| | (3.244) | (3.225) | (0.025) |
| Panel B: College entrance exam scores (standardized by year $\times$ track) | | | |
| Total | 0.182 | -0.586 | 0.767*** |
| | (0.939) | (1.047) | (0.000) |
| Chinese | 0.137 | -0.437 | 0.574*** |
| | (0.972) | (1.002) | (0.000) |
| Math | 0.165 | -0.534 | 0.698*** |
| | (0.945) | (1.035) | (0.000) |
| English | 0.136 | -0.447 | 0.583*** |
| | (0.974) | (0.997) | (0.000) |
| Comprehensive (STEM) | 0.181 | -0.826 | 1.007*** |
| | (0.955) | (0.950) | (0.000) |
| Comprehensive (Non-STEM) | 0.082 | -0.222 | 0.304*** |
| | (0.956) | (1.021) | (0.000) |
| Panel C: College admission | | | |
| Any college | 0.674 | 0.248 | 0.426*** |
| | (0.469) | (0.432) | (0.000) |
| Elite college | 0.016 | 0.002 | 0.014*** |
| | (0.124) | (0.043) | (0.000) |

*Notes:* This table presents summary statistics for students attending high schools that do and do not serve as college entrance exam testing centers. Each cell reports the mean, with the standard deviation in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[Back to Page 26]

## Appendix B Discussion of Major Choice

In this section, we examine whether the decline in STEM performance at a non-home test center is associated with students' subsequent choices of college type or field of study (in particular, the likelihood of selecting a STEM major).

Some colleges are more technically oriented, such as the University of Science and Technology of China. We define whether a college is technically oriented based on its name. Specifically, if the name contains keywords such as "Science and Technology," "Engineering," "Medical," "Polytechnic," "Agricultural," "Pharmaceutical," "Architecture," or "Institute of Technology," we classify it as technically oriented.

In addition to selecting universities, students in China must also choose specific majors when completing their college applications. Most majors are categorized by academic track and are typically available only to either liberal arts or science students. For example, majors such as physics, engineering, and computer science are generally restricted to science-track students, while fields like Chinese literature and history are limited to those on the liberal arts track. Some interdisciplinary or broadly defined programs, such as economics or law, accept applicants from both tracks. In other words, whether a student pursues a STEM or non-STEM major in university is largely determined by their choice of the science or liberal arts track after their first year of high school (Grade 10). However, science-track students have some flexibility to pursue non-STEM majors in college.

The results are reported in Panel A of Table A3. We do not find strong evidence that taking the exam at a non-home school reduces the likelihood of choosing a technically-oriented college or a STEM major.

## Appendix C Benefit-Cost Analysis

This section presents a back-of-the-envelope calculation for the net social benefit (cost) of a policy that assigns every student to a home-school test center in our context. That is, what if each student sits the NCEE at their own high school, relative to the status quo under which the majority of students are routed to off-site venues. We adopt a partial-equilibrium framework that accounts for three categories of benefit: (i) direct cost savings from avoided exam retakes, (ii) long-run wage gains from eliminating the earning penalty documented in Section 4, and (iii) improved geographic mobility, which we conservatively omit from the central estimate. On the cost side, we draw on administrative expenditure records from our sample county to construct a bottom-up accounting of both fixed capital outlays and annual operating costs for a standard, 900-student test center.

**Costs.** Panel A of Table C1 presents the full cost accounting. Fixed capital expenditures — covering surveillance infrastructure, signal-shielding equipment, the acoustic broadcast system, and the secure document room — total ¥900,000 at first installation. We annualize these outlays using asset-specific straight-line depreciation schedules ranging from 8 to 20 years, yielding an annual capital charge of ¥80,000. Annual operating costs amount to ¥237,000, dominated by invigilator compensation (¥108,000), paper-transport and security logistics (¥40,000), and emergency power provision (¥30,000). The total annual cost of operating one test center is therefore C = ¥317,000, or approximately ¥352 per student served.

**Direct Benefits.** Panel B.1 of Table C1 captures the private cost savings from avoided exam retakes. Our causal estimate (Table 1, Column 8) indicates that assignment to a non-home center raises the probability of retaking the NCEE by 13.3 percentage points. We use this retake probability as the relevant weight, since the direct costs at stake (private preparatory school tuition and foregone labor income) are incurred specifically by students who retake, not by all students who miss college in a given year. The expected retake costs per student are ¥1,995 in tuition (= 13.3% × ¥15,000) and ¥6,384 in foregone income (= 13.3% × ¥48,000), together with ¥600 in travel and accommodation expenses. Summing these components yields an expected direct benefit of ¥8,979 per student. Aggregated across the 900-student cohort, total annual direct benefits amount to ¥8,081,100.

**Wage Benefits.** Panel B.2 of Table C1 monetizes the earnings penalty estimated in Column 9 of Table 1. We find that students assigned to off-site centers earn approximately 2 percent less upon entering the labor market. Applying this coefficient to a baseline annual

**Table C1: Benefit-Cost Analysis of Home-School Test Center Policy**

| Category / Item | Annual Value (¥) | Per Student (¥) | Parameter Source |
|---|---|---|---|
| **Panel A: Costs** | | | |
| *A.1 Fixed Costs (Annualized Capital Expenditure)* | | | |
| Video surveillance system | 25,000 | 27.8 | ¥250,000 / 10 yrs |
| Signal-shielding & detection equipment | 25,000 | 27.8 | ¥200,000 / 8 yrs |
| Listening/broadcast system (backup) | 15,000 | 16.7 | ¥150,000 / 10 yrs |
| Secure examination room construction | 15,000 | 16.7 | ¥300,000 / 20 yrs |
| Subtotal — Fixed | 80,000 | 88.9 | $\Sigma$ depreciation |
| *A.2 Variable Costs (Annual Operating Expenditure)* | | | |
| Exam paper transport & security escort | 40,000 | 44.4 | Admin records |
| Office supplies & disinfection | 15,000 | 16.7 | Admin records |
| Emergency power backup & inspection | 30,000 | 33.3 | Admin records |
| Radio-frequency monitoring patrol | 10,000 | 11.1 | Admin records |
| Traffic control & environmental management | 25,000 | 27.8 | Admin records |
| Invigilator stipends (60 staff $\times$ 3 days) | 72,000 | 80.0 | ¥400/person-day |
| Invigilator travel & accommodation | 36,000 | 40.0 | ¥200/person-day |
| Local security & admin staff supplement | 9,000 | 10.0 | ¥150/person-day |
| Subtotal — Variable | 237,000 | 263.3 | |
| **Total Annual Cost (C)** | **317,000** | **352.2** | A.1 + A.2 |
| **Total Annual Cost (C) in USD** | **45,863** | **51.0** | A.1 + A.2 |
| **Panel B: Benefits** | | | |
| *B.1 Direct Benefits: Avoided Retake Costs* | | | |
| Expected retake probability | 13.3% | – | Table 1, Col. 8 |
| Private retake school tuition (avoided) | 1,795,500 | 1,995.0 | ¥15,000/retaker |
| Foregone income during retake year | 5,745,600 | 6,384.0 | ¥48,000/retaker |
| Transport & accommodation | 540,000 | 600.0 | ¥600/retaker |
| Subtotal — Direct Benefits | 8,081,100 | 8,979.0 | |
| *B.2 Long-Run Wage Benefits (Annual Equivalent)* | | | |
| Students benefiting | 900 | – | Full cohort |
| Wage penalty avoided | 2% | – | Table 1, Col. 9 |
| Baseline annual wage | 48,000 | – | Admin data |
| Annual wage gain per student | – | 960.0 | ¥48,000 $\times$ 2% |
| Subtotal — Wage Benefit | 864,000 | 960.0 | |
| **Total Annual Benefit (B)** | **8,945,100** | **9,939.0** | B.1 + B.2 |
| **Total Annual Benefit (B) in USD** | **1,294,169** | **1,438.0** | B.1 + B.2 |
| **Panel C: Summary Statistics** | | | |
| Net Annual Benefit (B − C) | 8,628,100 | 2,850.4 | |
| Net Annual Benefit (B − C) in USD | 1,248,306 | 412.0 | |
| Benefit-Cost Ratio (B / C) | 28.2 | – | |
| Students served per center | 900 | – | |
| Break-even cost threshold (¥ / student) | – | 9,939.0 | $= B/N$ |

*Notes:* Cost parameters are drawn from administrative records of the sample county's 2016–2018 NCEE administration. Benefit parameters are based on causal estimates reported in Tables 1 and 2. Wage benefits assume a 35-year career, 3% discount rate, 4% real wage growth, and baseline wage of ¥48,000. Direct benefits are computed using the 13.3 percentage-point increase in retake probability multiplied by per-retaker costs. One U.S. dollar (USD) is approximately 6.9 Chinese yuan (CNY).

starting wage of ¥48,000 implies an annual wage gain of ¥960 per student. Summed across the full cohort of 900 students, the aggregate annual wage benefit is ¥864,000. A full present-value calculation, assuming a 4% real wage growth rate, a 3% social discount rate, and a 35-year career, yields a per-student NPV of ¥38,628 — consistent with the estimate reported in our companion cost-benefit table. For comparability with the annual cost measure, we report the annualized equivalent in Panel C.

**Net Benefit and Benefit-Cost Ratio.** Summing across both categories, the total annual benefit is B = ¥8,945,100. Against a total annual cost of ¥317,000, the net annual benefit is ¥8,628,100, yielding a benefit-cost ratio (BCR) of B/C = 28.2. Put differently, each dollar invested in home-school testing infrastructure returns approximately 28 dollar in measurable social value. The break-even cost threshold, the maximum per-student cost at which the policy would still be justifiable on cost-benefit grounds, is ¥9,939. This is more than 28 times the actual per-student cost of ¥352. This large margin of safety implies that the positive net-benefit conclusion is robust to substantial upward revisions in cost estimates or downward revisions in the causal estimates, and it holds even after entirely excluding the wage-benefit channel (BCR = 25.5 under direct benefits alone).

**Limitations and Scope.** Several caveats apply. First, our cost estimates reflect administrative data from a single county and may not generalize to settings with different infrastructure baselines or labor costs. Second, the wage benefit is estimated from a reduced-form coefficient and implicitly assumes that the entire earnings gap is attributable to the assessment-location shock rather than to the composition of colleges attended; robustness checks reported in Table A6 support this interpretation but cannot entirely rule out residual confounding. Third, we omit several plausible benefit channels — including the value of improved geographic mobility (Column 10 of Table 1) and reduced psychological distress among students — which would, if quantified, further strengthen the case for the policy. We therefore regard our BCR estimate of 28.2 as a lower bound on the true social return.