Social Network Formation and Exam Fraud: *A Friend in Need is a Friend Indeed*

Tam Mai[†]

Job Market Paper

November 20, 2024

Click here for the latest version

Abstract

The proliferation of standardized testing has raised concerns about its distortionary effects on both school and student behavior. While previous literature has extensively documented high-powered misconduct by teachers and administrators, there is little systematic evidence on how exam pressure may lead students to commit academic dishonesty. This paper investigates interpersonal student cheating during high school exit exams in Vietnam, a setting where educational success is highly coveted. Using individual-level data from a large province, I leverage the quasi-random assignment of students to test rooms to estimate peer effects on test day. I find that students from low-ranked schools performed better when taking exams with students from elite schools. High-achieving elite peers are particularly valuable. However, the gains were concentrated in multiple-choice and quantitative tests, but absent in essay exams. Moreover, the positive effects virtually disappeared after a testing overhaul increased the stakes of the exams. Backed by institutional details, these suspicious patterns provide credible evidence that discreet interpersonal cheating was once prevalent: non-elite students quickly formed networks with elite peers to cheat for their own benefit. It took a major reform to reshape student incentives and eliminate this malpractice. (*JEL* I21, I28)

[†]Department of Economics, Columbia University and Norwegian University of Science and Technology (NTNU); Email: tam.mai.vii@gmail.com. I am indebted to Sandra Black, Jonas Hjort, and Miguel Urquiola for their continuous guidance and support. I also thank Michael Best, Colin Green, Andreas Kostøl, Cristian Pop-Eleches, Jonah Rockoff, Hammad Shaikh, Rodrigo Soares, seminar participants at Columbia University, NTNU, and Drexel University, and the audience at the 2023 Workshop on Education Economics and Policy for helpful discussions and comments. All errors are my own. Financial support from Columbia University's Center for Development Economics and Policy (CDEP) is gratefully acknowledged.

Standardized tests are ubiquitous: "no one knows with certainty how many [...] are being given every year, what proportion of the population is being or has been tested, or precisely what use is made of the scores that result" (Goslin 1963, p. 45). These exams, a common feature in many school accountability systems, are primarily used to evaluate student competence and school quality. As testing became a "culture" (Moses and Nanna 2007), however, the controversy surrounding its use has only intensified. Critics argue that excessive emphasis on testing can distort both school and student behavior (Linn 1991). In addition, rather than measure students' inherent ability, test outcomes may reflect unequal access to resources, such as test prep, thereby perpetuating inequality. It is thus unclear whether standardized tests are the best solution to the question of assessing educational quality (Hanushek and Raymond 2003).

While the informativeness of standardized tests is widely debated, less attention has been paid to the integrity of *test scores*. It is customary for empirical studies to use standardized test scores as proxies for student ability or accomplishment (Jacob and Rothstein 2016). Despite test design imperfections, grading is believed to be honest and free from interference: the marks at least capture true mastery of the tested material. However, this interpretation becomes problematic in light of manipulations like grade inflation. An obvious, albeit risky, way to inflate test achievements is outright cheating. Cheating can be grassroots, with students, the immediate beneficiaries, inflating their own scores. It can also be high-powered, where administrators and teachers are at fault (Jacob and Levitt 2003). Finally, cheating can be a combination of both bottom-up and top-down misconduct, in which both students and educators are complicit (Borcan, Lindahl, and Mitrut 2017). In this paper, I consider a variant of this "hybrid" fraud: students take the initiative to cheat on exams, but possibly with tacit aid from negligent proctors.

The setting is the National High School Exams (NHSE) in Vietnam. This country is considered an outlier in education: despite its low-middle-income status, Vietnam often outperforms wealthier nations in international student assessments (Dang et al. 2023). Educational success is highly valued in this culture. Rigorous standardized testing for talent allocation dates back hundreds of years, to the feudal era when Confucian teachings dominated (Stankov 2010, Marginson 2011). Traditional Confucian beliefs in the importance of education continue to prevail today as testing is modernized. One of the milestones in present-day student life is the NHSE. Up until 2015, this exam was used to graduate students from high school and screen candidates for the National College Admission Exams (NCAE). In 2015, the two assessments were consolidated to serve both high school graduation and college admission purposes. Although cheating was reportedly widespread before the reform (Kiều 2012,

H. 2014), the overhaul made the NHSE more competitive and changed student incentives, likely reducing the appeal of cheating.

Using the NHSE as a laboratory, this paper aims to detect mass student cheating on exam day. To this end, I ask two questions. First, do students from non-elite schools perform better if they happen to take the exam alongside students from elite schools? Second, how do these effects evolve in relation to the timing of the reform? A non-elite student may befriend an elite student in their room during the exam window so as to copy their exam answers (active cheating on the non-elite student's part) or the elite peer can reveal their answers to them (passive cheating on the elite student's part). The giver is arguably more compassionate and altruistic when the exam is not competitive. This suggests that student cheating might have lessened after the reform.

The mixed-stakes nature of the NHSE makes it well-suited for studying this form of cheating. For the college-bound, high-ability students, succeeding in the subsequent NCAE is the ultimate goal. The learning material that these students must master to excel on the NCAE is much more demanding than that for the NHSE. In contrast, for the non-college-hopefuls, the NHSE can be challenging. In addition, the stakes of the NHSE within each group of students has varied over time, not only with the implementation of the 2015 reform, but also with the intensity of government efforts to curtail cheating. The nature of the NHSE is therefore ambiguous: low-stakes for high-ability students, but high-stakes for low-ability students; overall low-stakes in some years, but high-stakes in others.

Empirically, I use individual test score data from a large province in Vietnam that cover two periods: pre-reform (2007-2013) and post-reform (2018-2019).[1] Apart from test scores, the data contain detailed information on the location where each student took the NHSE, the name of their high school and classroom, as well as their basic demographic and socioeconomic characteristics. For identification, I leverage the quasi-random assignment of students from schools of different quality into test rooms. This quasi-randomness stems from the alphabetization of student names for test administration purposes. On a subset of the data, I also harness information on the elite student's high school classroom name, which signals their subject and field of specialization, in an instrumental-variables (IV) framework to estimate the contemporaneous effects of specialized students' performance on non-specialized peers'—the very effects that this paper will argue are indicative of cheating.

I find that, before the reform, non-elite students benefited from having immediate access to elite peers, as measured by the presence or the share of elite students in their test room.

1. Data are missing or incomplete from 2015-2017, precluding the use of these years in the analysis.

However, the gains were not uniform across subjects. No effects were detected in the essay, open-response tests (Vietnamese, History, and Geography). For the multiple-choice or quantitative subjects (Physics, Chemistry, Biology, Foreign Language, and Math), in contrast, sharing the same test room with at least one elite student increased non-elite students' test scores by 0.1-0.2 standard deviations (sd). Moreover, the effects of same-room *elite density* were largely constant across its distribution: depending on the subject, a 10 percentage point (pp) increase in the share of elite students in a test room raised the test scores of non-elite peers by 0.02-0.04 sd. In the wake of the 2015 reform, however, these positive effects virtually vanished.

Similar patterns emerge in the IV subsample, where I use the average test scores of elite students in a test room to proxy for *elite quality* and quantify its impacts on non-elite roommates. Elite peers of higher quality were instrumental, but only before the reform and in the multiple-choice subjects: one sd higher in the average quality of elite roommates translated into a 0.4-0.5 sd increase in non-elite students' scores in Physics, Biology, and Foreign Language. The effects of both the quantity and quality of elite peers varied across levels of student ability, but exhibited little heterogeneity by gender. All in all, the asymmetries across subject types and time periods lend credence to the argument that the positive impacts observed prior to the exam redesign reflected cheating rather than benign peer effects.

Related Literature This paper nests within the broad literature on the intended and unintended consequences of test-based incentives. Such incentives have been found to be effective in boosting test scores in the short term (Grissmer et al. 2000, Carnoy and Loeb 2002, Hanushek and Raymond 2003, Hanushek and Raymond 2005, Jacob 2005, Angrist and Lavy 2009, Muralidharan and Sundararaman 2011) and sometimes effective in increasing educational attainment in the long term (Deming et al. 2016). At the same time, test-based incentives can distort behavior. For example, teachers may teach to the test (Jacob 2005, Glewwe, Ilias, and Kremer 2010) and schools may over-invest resources in high-stakes subjects at the expense of others or strategically re-engineer the test pool (Jacob 2005, Cullen and Reback 2006, Figlio 2006, Figlio and Getzler 2006). These hidden behaviors are often only brought to light by "forensic economics" (Zitzewitz 2012).

Few economic studies, however, focus on outright cheating as a response to exam pressure. One notable exception is Jacob and Levitt 2003. Using standardized test score data from Chicago's public schools system, the authors develop an algorithm to identify probable instances where teachers or school administrators have systematically tampered with answer sheets after exams to inflate student scores. The algorithm accounts for both abnormal within-student fluctuations in test scores across years and unusual patterns of answers on

the exam sheets of different students in the same classroom. This analysis unearths probable cheating in about five percent of Chicago's elementary school classrooms every year between 1993 and 2000.

In another related study, Borcan, Lindahl, and Mitrut 2017 examine the distributional effects of an anti-corruption campaign targeting fraud in the Romanian Baccalaureate. The campaign reinforces punishment for cheating and strengthens monitoring through the installation of closed-circuit television (CCTV) cameras at test centers. Exploiting the staggered introduction of CCTV surveillance across counties, the authors find that the two measures both succeeded individually and complemented each other in curbing test fraud. However, the well-intentioned initiative widened the score gap between rich and poor students. The authors hypothesize that the campaign might have encouraged rich, well-connected students to step up bribes to examiners. Poor students, being less connected and financially unable to match those outsize gifts, were outbid. As CCTV was unable to catch this *quid pro quo*, the campaign ended up hurting socioeconomically disadvantaged students.

The present paper contributes to this sparse literature in several dimensions. First, in contrast to Jacob and Levitt 2003 who discover after-the-fact, top-down cheating by teachers and school administrators, I focus on real-time, bottom-up cheating by students. Like this prior study, my analysis goes beyond investigating sporadic cheating incidents that are often too severe to escape suspicion. Instead, I identify systemic, surreptitious misconduct and examine how this behavior responds to changes in incentives. Furthermore, while cheating in general implies some sort of conflict between the involved parties, the type of cheating under scrutiny requires cooperation between students. Due to the particular ways in which the NHSE is designed and implemented, students are willing to coordinate despite the fact that the benefits are privately accrued. In other words, there can be winners without losers.

Second, this paper offers new evidence of exam cheating in a unique setting: a developing country that, despite poor economic status, has paradoxically managed to pull above its weight in education. Like Romania studied in Borcan, Lindahl, and Mitrut 2017, Vietnam is plagued by corruption. Policy-makers in Vietnam are not oblivious to the egregious corruption in testing and have vowed to curb it. Unlike in Romania, however, official efforts to confront cheating in Vietnam are often short-lived. Monitoring remains lax and ill-equipped while punishment is neither strictly enforced nor harsh enough to be an adequate deterrent. I thus explore how subtle forms of cheating persist in a setting without high-tech surveillance and sustained monitoring efforts.

Last but not least, my work is also linked to media coverage of exam cheating scandals.

Reportage of testing misconduct in the NHSE has often attracted public attention and catalyzed test reforms (Kiều 2012, H. 2014, Nguyễn 2016). To make headlines, however, cheating must be unmistakably glaring. Examples of such audacity, shown in Appendix Figure B1, include unauthorized use of notes or mobile devices connected to helpers-in-waiting out of the test location (Panel A). Exam questions are leaked outside and solutions are delivered to students afterward (Panel B). Meanwhile, less blatant forms of dishonesty, such as furtive transmission and quiet copying of answers inside the test room (Panel C), slip under the radar. This paper zooms in on one such covert fraud: in-room cheating between students from low- and high-ranked schools.

The remainder of the paper is organized as follows. Section 2 provides institutional details about education in Vietnam and the NHSE. Section 3 describes the data and variable construction and provides empirical support for the main identification strategy. Section 4 estimates the effects of the quantity of elite students on the test scores of non-elite peers; Section 5 does the same for the quality of elite students. Section 6 discusses the estimation results, crystalizing the evidence of cheating, and briefly concludes.

2 Background

2.1 Education in Vietnam

Pre-tertiary education in Vietnam consists of four levels: kindergarten, elementary school (grades 1-5), middle school (6-8), and high school (10-12). Public and private schools co-exist at all levels, but public schools outnumber and enroll more students than private schools. Admissions to public middle and high school are competitive and meritocratic: both require taking entrance exams at the district and province level, respectively. Exam eligibility is based on local residency and admission thresholds vary across schools. In particular, only students whose test scores exceed a school's score cut-off are admitted.

At the high school level, two main programs exist: academic, offered in formal high schools, and vocational, available in institutions like continuing education centers and trade schools. Unlike formal high schools, the latter institutions mainly serve adult learners or students with interrupted educational history. Admission to these schools is non-competitive, with open enrollment being the norm. Within each program, students are further tracked into either physical science or social science and humanities.

Each province operates at least one specialized public high school that is overseen by the provincial Department of Education and Training (DOET). Public universities located in

the province may also operate their own specialized high school programs outside the local DOET's purview. Admission to specialized high schools is separate from that to regular schools and involves extra entrance tests. All applicants take common exams in two core subjects (Math and Vietnamese, cf. SAT I) and an additional exam in the specialized subject of their choice (cf. SAT II). The final score used for admission considerations is a weighted average of the three test scores, with a higher weight on the specialized subject.

Beyond high school, college admission is also contingent on competitive exams, but at the national level: the NCAE before 2015 and the NHSE since. Test subjects are grouped into blocks of three. For example, Math, Physics, and Chemistry constitute block A_1 whereas Math, Vietnamese, and Foreign Language make up block D_1, and so on. Universities admit students by block of subjects, with different score cut-offs for different blocks. Unlike pre-tertiary education, college application does not require local residency or household registration, although some preferential treatment is given on the grounds of socioeconomic status such as ethnic minority and special family circumstances.

An interesting feature of college admission in Vietnam plays an important role in my analysis: not every student is eligible for the college entrance exams. As a prerequisite, college applicants must have successfully graduated from high school first. Pre-reform, this amounted to passing the stand-alone NHSE where all students were tested in the same predetermined subjects. From 2015 on, the composite NHSE allows for customization: students can take up to eight subject tests, six of which (Math, Vietnamese, Foreign Language, and three electives) count toward high school graduation. In both cases, a pass grade is granted if a student scores at least an average of five across the relevant subjects. The next section provides more detail.

2.2 The NHSE

The following description is based on the official documentation of the NHSE regulations (Ministry of Education and Training of Vietnam 2007).

Overview and History The NHSE is an annual centralized high school exit exam. Pre-reform, grade 12 students from different high school tracks (academic or vocational) were evaluated separately. Since 2015, this distinction has been abolished and all examinees take the same and only exam track. The assessment takes place at the end of May or early June after the school year ends. Students who fail the exams in a given year will generally need to wait for at least a year for a re-take. On rare occasions (e.g., in 2007 and 2008), re-takes were available in the same year, but only after that year's NCAE.

A minimum number of six subject tests are needed for high school graduation. Math and Vietnamese are mandatory. Depending on a student's high school track and a high school's resources, Foreign Language is either required or can be substituted with another subject on a case-by-case basis. Pre-reform, the Ministry of Education and Training (MOET) dictated the remaining three subjects. Starting in 2015, the number of additional subjects is raised to six, including a brand new test subject (Political Science). Students can select three electives by field: physical science (Physics, Chemistry, and Biology) or social science and humanities (History, Geography, and Political Science).

As summarized in Table 1, the test format differed across subjects and periods. Until 2014, half of the tests were multiple-choice (Physics, Chemistry, Biology, and Foreign Language), while the others were open-response (Vietnamese, Math, History, and Geography). Within a high school track (pre-2015) and across both tracks (since 2015), there is one test booklet per essay exam, but different test booklets for each multiple-choice exam. However, the differences are cosmetic. The multiple test booklets only differ in the order of test questions and answer options; the actual test contents are identical. Next, I briefly discuss the evolution in the competitiveness of the NHSE over the last two decades.

TABLE 1 Subject Test Format

	Pre-reform		Post-reform	
	Multiple-choice	Write-in	Multiple-choice	Write-in
Physics	×		×	
Chemistry	×		×	
Biology	×		×	
Foreign Language	×		×	
Math		×	×	
Vietnamese		×		×
Geography		×	×	
History		×	×	
Political Science	-	-	×	

Pre-reform NHSE The NHSE's foremost purpose was to graduate students from high school. On the one hand, since a student's performance on the NHSE determined whether they were eligible for the NCAE, the NHSE was not just a formality. On the other hand, the NHSE was not as high-stakes as the Romanian Baccalaureate studied by Borcan, Lindahl, and Mitrut 2017. While also a high school exit exam, the Romanian Baccalaureate is much more critical to college application: university admission is based in large part on student scores on this exam. In contrast, a student's performance on the Vietnamese NHSE

only mattered to the extent that a pass was needed for the student to qualify for the NCAE. For college admission, a score higher than the passing grade on the NHSE provided no additional advantages.

At the same time, the NHSE was not equally challenging to all students. For the college-hopeful, who tend to be higher-achieving, succeeding in the NCAE was the ultimate goal. The study material that they had to master to succeed on the NCAE was more advanced than that for the NHSE. For other students, the NHSE was not necessarily a cakewalk. Consequently, prior to the reform, the nature of the NHSE was ambiguous: low-stakes for high-achievers, but high-stakes for low-achievers.

Post-reform NHSE Not only do the stakes of the NHSE vary across groups of students, but they also change over time within each group. Most notably, the 2015 reform streamlined testing and combined the NHSE and the NCAE into one single assessment. Moreover, students from different high school programs are no longer separated into different exam tracks for evaluation purposes: everyone now receives the same exams. This consolidation makes it harder for students to discern other test-takers' objectives, and so, possibly increases perceived competition. As more is at stakes, the government also doubles down on its efforts to curb cheating. These changes in student incentives and monitoring efforts disambiguate the nature of the NHSE: following the reform, the exam has undoubtedly become competitive.[2]

Given the history and design of the NHSE, two questions emerge. First, could cheating among students of different high school calibers have occurred in the pre-reform period, despite the government's surveillance efforts? Second, has this interpersonal form of cheating persisted into the post-reform period? To provide intuition, I formalize the problem of interpersonal cheating in a stylized model presented in Appendix A and derive predictions to inform the empirical analysis. The main empirical objective is to identify peer effects on test day and their evolution, using both cross-sectional and temporal differences in exam stakes. Crucial to my identification strategy is how the NHSE is administered in practice.

Test Administration While the NHSE is a centralized assessment, with the same test

2. Apart from the major reform in 2015, there have been sporadic smaller-scale initiatives that aim to combat test fraud head-on. For instance, in 2006, a whistle-blower in the province of Hà Tây reported rampant cheating at the test center where he was a proctor. The scandal sent shock waves across the country and galvanized policy-makers into finding immediate solutions. The impact of the increased efforts to combat cheating was felt at once. In 2007 and 2008, the national share of first-attempt passes nearly halved from a high of 90% in the previous years (Kiều 2012). This nosedive prompted the administration of a re-take within two months in both years—a heretofore unprecedented decision. Unfortunately, as the costs of containing cheating became prohibitive, efforts waned and the old status quo of inflated passing rates was instantly restored.

questions and grading rubrics nationwide, the administration of the exam is localized. Each province's DOET determines the test locations where local students take the NHSE. Almost always, students are to take the exam in a test center in their home district. Under rare circumstances, some students might be assigned to a test center in a nearby district if this test center is closer to their high school than those in their home district. A test center is a local high school or middle school campus. However, a student's test center may not be the same as their high school building. The exact test center depends on which test committee the student's high school belongs to.

The configuration of test committees varies from year to year. Oftentimes, there are multiple test committees within a district, each consisting of several high schools. The exact groupings depend on geographic proximity. In some cases, an entire district is designated as a single test committee, so all grade 12 students in the district are completely mingled. Very seldom, each committee corresponds to a single high school. In all cases, after the list of test committees is finalized, each committee is assigned to a test center, namely the physical location where the students of the member schools will sit for the NHSE.

The next step is to assign students to test rooms of at most 24 seats each. As illustrated Figure 1, the test rooms are regular classrooms with a standard capacity of 12 long desks.[3]Within each test center, students are first sorted by the foreign language in which they will be tested (e.g., English, French, or Russian). Within each resulting stratum, all student names are ordered alphabetically: first by first name, then by last name, and finally by middle name. The students are then assigned to different test rooms according to this order.

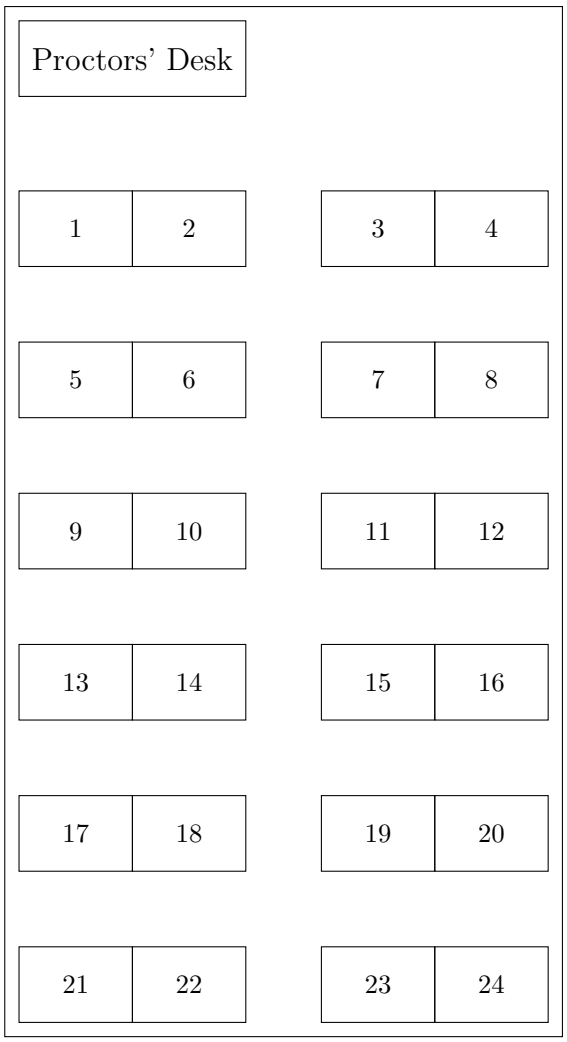
The above alphabetization of student names creates useful exogenous variation in student composition across test rooms. It is worth noting that unlike person names in other cultures, Vietnamese person names generally do not signal socioeconomic status.[4] This is particularly true for first names, which are the first variable used to sort students for test room allocation. Moreover, as will be shown in Section 3, in this context, a student's name is unlikely to be correlated with their cognitive ability and the quality of their high school. Within each test

3. The 12 desks are arranged into two columns of six desks, with each desk seating two test-takers. Seat arrangements must satisfy two conditions: first, the distance between the two students at the same desk must be at least 1.2 meters, and second, students sitting immediately in front of or behind each other must be perfectly aligned. Proctors have some leeway to assign students to seats as they see fit. (Seating arrangements are not observed in the data.) Moreover, after each test session, the seats are reshuffled, ensuring that no seating chart is repeated.

4. Some names and spellings can signal that the name-holders are of ethnic minority. Nonetheless, Vietnam is rather ethnically homogeneous: about 90% of the population are of the majority ethnicity, Kinh. For this reason, the vast majority of Vietnamese person names do not have an ethnic minority connotation.

center and foreign language stratum, the assignment of students into test rooms is therefore plausibly quasi-random.[5]

FIGURE 1 Test Room Seating Chart



Notes: This diagram illustrates a typical seating chart in a test room with 12 desks arranged in two columns of six. A desk consists of two unenclosed cells, each accommodating one student. The minimum distance between any two students at the same desk is 1.2 meters. To fix ideas, the numbers shown in the diagram index the students in the test room. In practice, the proctors can arrange the students in any order, although this information is not recorded in the data.

5. In this case, the alphabetization of student names is unlikely to be susceptible to Deaton 2010's (2010) critique of the misuse of alphabetization for identification. Deaton 2010 questions whether alphabetization of school names, as in Miguel and Kremer 2004, yields true quasi-randomness. In my setting, however, the alphabetization happens at the student level and affects all students. Furthermore, it is implemented only for the purpose of assigning test ID numbers and test rooms. It is hard to conceive how this alphabetization may still leave confounders lurking, especially in light of how uninformative Vietnamese person names are.

Such quasi-randomness is critical to obtaining causality, but inadequate on its own. I conjecture that, prior to the reform, cheating was present in the non-competitive NHSE: students of lower-ranked high schools copied from students of elite high schools. This hypothesis would be untestable without the following institutional detail. The full roster of students by test room was published at the entrance to each test center. The list of students in each test room was also posted on the room's front door. Each list included personally identifiable information such as a student's test ID number, their full name and date of birth, as well as their *high school name*. In addition, before each test session, students were asked to queue outside in the hallway. No one was permitted to enter the test room until a proctor called their name, checked their ID, and directed them to their seat. It is because of this high level of "transparency" that students in the same test room could readily identify one another. Along with exogenous variation in student composition across test rooms, this transparency allows me to identify cheating as hypothesized.

3 Data Construction and Descriptive Statistics

3.1 Data Source and Sample Restrictions

My analysis uses individual-level test scores data from a large province in Vietnam. Table 2 gives an overview of the student population in this province. There are 21 school districts with around 90 (110) high schools with NHSE test-takers before (after) the 2015 reform. After 2015, there are more NHSE schools because the reform unified the academic and vocational programs into a single exam track. The higher number of test schools after 2015 includes vocational schools that newly participate in the unified NHSE. The total number of test-takers also varies slightly over time. From 2007 to 2019, about 30-45,000 students in the province took the NHSE annually. The capital city was the largest school district, with 12-15 high schools (two of which are specialized) and about 4,000-5,000 test-takers (over 700 of whom were specialized students) each year in the data.

The full test scores data are available from 2002-2019. However, I restrict my main analysis to seven student cohorts pre-reform (2007-2013) and two cohorts post-reform (2018-2019), dropping the intervening years due to inadequate data. From 2014-2017, either detailed test locations are missing or a typical test room featured too few schools to generate sufficient variation in student composition by elite status. On the one hand, if cheating is more likely within schools, mixing students from different schools helps restrain the behavior. On the other hand, this also creates opportunities for students from low-quality schools to seek help from higher-achieving peers. Given my objective of detecting this very type of cheating, it is

TABLE 2 Basic Description

	Pre-reform							Post-reform	
	2007	2008	2009	2010	2011	2012	2013	2018	2019
No. school districts	21	21	21	21	21	21	21	21	21
No. high schools (HS)	86	90	91	91	91	91	91	111	109
No. elite HS	22	22	22	22	22	22	22	22	22
No. specialized HS	2	2	2	2	2	2	2	2	2
No. test-takers	45,072	48,836	43,118	39,418	40,576	38,748	38,412	29,655	30,966
No. elite HS test-takers	12,139	13,016	11,644	10,345	11,127	10,855	11,006	9,039	9,225
No. HS in capital city (CC)	12	12	12	12	12	12	12	15	15
No. test-takers in CC	4,795	4,823	4,963	4,771	4,893	4,273	4,346	3,446	3,613
No. specialized HS test-takers in CC	658	743	804	754	964	826	807	812	810

Notes: This table tallies and categorizes the schools and students that took the NHSE in the years of interest.

appropriate to focus on 2007-2013 and 2018-2019.

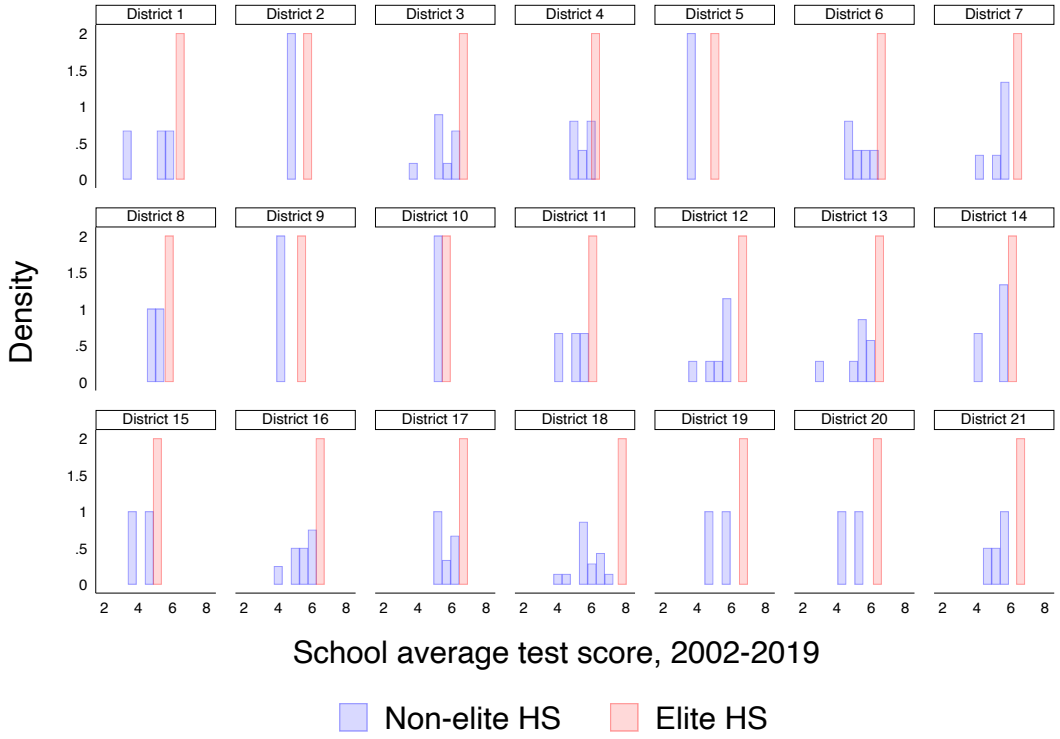
3.2 Variable Construction

Elite Status A student is classified as elite if they go to an elite high school. Given their low acceptance rates, both specialized high schools in the province are considered elite schools. Apart from these magnet schools, school eliteness is defined by district. Within a district, the elite school is designated as the high school with the highest average composite NHSE score across all enrollees between 2002 and 2019. In the capital city, this coincides with one of the two specialized high schools. Hence, there is no loss in classifying the specialized high schools as elite.

Although the main estimation is based on fewer years of data, I use all test scores from 2002-2019 to define elite schools for several reasons. First, school average test scores may be sensitive to measurement error due to differences in enrollment across schools (Kane and Staiger 2002). With more school-level observations spread out over a longer period, this measurement error is smaller, bolstering the categorization of eliteness on the basis of school average test scores. Second, as a robustness check, I also consider a second algorithm that, within a district and given a year, assigns elite status to the high school with the highest *moving average* of composite NHSE scores. Unlike the first method, this approach can theoretically lead to in time-varying elite status: a school might be classified as elite in one year only to lose this status later. In practice, however, the two algorithms agree almost everywhere. On the few occasions where they deviate, careful inspection and consultation with the DOET of the province in question reveal that the assignment of eliteness by the

first method is more reliable.[6] As confirmed by the near perfect overlap in the output of the two algorithms, school reputation is highly persistent. Thus, the rest of the paper uses the time-invariant elite status determined by the first method. Using this definition, Figure 2 illustrates the histogram of school-level test scores by school type. Within each district, the elite school(s) are distributed to the right of non-elite schools, providing internal validation.

FIGURE 2 Salience of Elite Status



Notes: This figure presents the histogram of school-level test scores by school type. The score for each school is the average composite score across all of its students who took the NHSE between 2002 and 2019.

Elite Quantity and Elite Quality For brevity, I refer to the quantity (quality) of elite students in a test room as elite quantity (quality). Elite quantity is measured in two ways: elite presence and elite quality. Elite presence is an indicator for the existence of at

6. The second algorithm is prone to outliers. Private schools typically enroll fewer students than public schools. Students of some small private school may happen to do better on average on the NHSE than students of large public schools in some year. With limited observations, the alternative algorithm classifies such a private school as elite. In reality, this could not be further from the truth: private schools charge higher tuition, do not admit students based on academic merits, and target leftover students, i.e., those who have failed to get into public schools. This misclassification error arises because unequal enrollments across schools exacerbate mismeasurement of school quality (Kane and Staiger 2002).

least one elite student in the test room, while elite density is calculated as the leave-one-out fraction of elite students present in the test room. For each room with at least one elite student, elite quality in a given test subject (e.g., Math) is computed as the average Math score across all elite students in that room.

3.3 Descriptive Statistics

Background Characteristics and Test Performance Table 3 summarizes student background characteristics and test performance by elite status, separately by period. In both the pre-reform (Panel A) and post-reform (Panel B) period, between a quarter and a third of the student population were elite. Compared to non-elite students, a greater proportion of elite students were female or ethnic minorities. Interestingly, before 2015, non-elite students were much less likely than elite students to be ranked as high-achieving within their school (Panel A), but the gap between the two groups vanished afterward (Panel B). As noted in Section 2, the grading standards that underly these rankings vary across schools. This catch-up by non-elite students likely reflects their response, along with schools, to changes in the formula for test outcomes: post-reform, not only exam scores but grade 12 GPA also matters.

Test performance on the NHSE both differed between groups of students and changed over time. First, elite students outperformed non-elite students in all subjects. Since the gross final score is simply the arithmetic mean of the subject scores and the number of the subject tests taken by both groups was identical, this score is also higher among the elite students. The net final score, the ultimate basis to determine pass/fail, is equal to the gross final score plus any bonus points due to the student's priority and ethnic minority status. As there were more minorities and prioritized students among the elite group, their net final score was again higher. Second, there were stark temporal variations in test performance. The average subject test scores were about 1-2 points lower in the post-reform period. Yet, a greater fraction of students passed the NHSE after (97%) than before the reform (81-86%), again likely because the reform introduced grade 12 GPA into the formula to determine the final test outcomes.

Test Room Personnel Characteristics Focusing on the group of non-elite students, Table 4 presents the descriptive statistics on student composition in their test rooms, including the number of high schools represented, a dummy for any elite students present, the number of elite students present, the fraction of elite students among other test-takers, and the average subject test scores of these elite students. The official cap of 24 test-takers per room was met almost always. Pre-reform, a typical test room mingled students from four

Table 3 Summary Statistics on Student Characteristics

A Pre-reform

	Non-elite			Elite		
	N	Mean	SD	*N*	Mean	SD
Background characteristics						
Female	214,048	0.52	0.50	80,132	0.54	0.50
Minority	214,048	0.05	0.21	80,132	0.14	0.34
No priority	214,048	0.82	0.38	80,132	0.71	0.46
Low priority	214,048	0.13	0.34	80,132	0.16	0.37
High priority	214,048	0.05	0.21	80,132	0.13	0.34
HS Academic standing: Low-achieving	214,048	0.63	0.48	80,132	0.49	0.50
HS Academic standing: High-achieving	214,048	0.37	0.48	80,132	0.51	0.50
Test performance						
Passing	214,048	0.81	0.39	80,132	0.86	0.34
Net final score	213,955	6.18	1.44	80,103	6.58	1.52
Gross final score	213,272	5.91	1.38	79,924	6.31	1.48
Physics	129,676	5.81	2.24	47,926	6.26	2.39
Chemistry	117,305	7.34	2.11	44,345	7.83	2.17
Biology	124,149	6.59	1.66	46,793	6.90	1.75
Foreign Language	214,044	4.96	1.87	80,130	5.47	2.07
Math	214,048	6.56	2.65	80,132	7.20	2.58
Vietnamese	214,048	5.42	1.49	80,132	5.63	1.54
Geography	145,295	5.96	1.24	54,977	6.16	1.25
History	125,719	5.09	2.17	46,355	5.40	2.20

B Post-reform

	Non-elite			Elite		
	N	Mean	SD	*N*	Mean	SD
Background characteristics						
Female	42,357	0.51	0.50	18,264	0.55	0.50
Minority	42,357	0.07	0.26	18,264	0.18	0.38
No priority	42,357	0.68	0.47	18,264	0.63	0.48
Low priority	42,357	0.24	0.43	18,264	0.19	0.39
High priority	42,357	0.07	0.26	18,264	0.18	0.38
HS Academic standing: Low-achieving	42,357	0.12	0.32	18,264	0.11	0.31
HS Academic standing: High-achieving	42,357	0.88	0.32	18,264	0.89	0.31
Test performance						
Passing	42,357	0.97	0.18	18,264	0.97	0.16
Net final score	42,275	5.08	1.42	18,263	5.51	1.01
Gross final score	39,831	5.03	0.92	18,262	5.23	1.04
Physics	12,660	4.97	1.74	7,214	5.43	1.86
Chemistry	12,660	4.95	1.68	7,214	5.33	1.81
Biology	12,660	4.38	1.24	7,214	4.42	1.36
Foreign Language	42,348	3.14	1.65	18,232	3.92	1.76
Math	42,357	4.87	1.68	18,264	5.34	1.88
Vietnamese	42,357	5.58	1.43	18,264	5.86	1.54
Geography	31,458	5.68	1.24	12,104	5.81	1.27
History	31,458	3.90	1.27	12,104	3.99	1.33

Notes: Panel A and Panel B summarize the pre-reform period (2007-2013) and the post-reform period (2018-2019), respectively. For each period, the sample consists of students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). All test scores are on a 0-10 grading scale.

high schools. On average, each test room had about six elite students, accounting for 28% of the room's capacity. There was substantial variation in the fraction of elite students across test rooms both in the cross section and over time. In particular, post-reform, a typical test room featured fewer (about three) high schools, but had more (seven to eight) elite students.

3.4 Evaluating Primary Identification Strategy

My approach to uncovering mass cheating by non-elite students hinges on both the quasi-random allocation of students into different test rooms within a test center and the observability of high school affiliation within a test room. Section 2 draws on institutional details to argue that both of these conditions hold in my context. This section complements Section 2 by providing direct evidence of exogeneity in test room assignment. In particular, I demonstrate that in my setting, the names of students predict neither the quality of their school nor their individual test performance.

Using data from all NHSE test-takers between 2002 and 2019, I first calculate the pairwise correlations between student names and academic performance. The results are reported in Table 5. All coefficients are close to zero, indicating no clear relationship between a student's name and either high school quality or test scores. Figure 3 and 4 illustrate this, showing binned scatterplots of elite school attendance and test performance by first name. A student's first name provides no insight into high school quality: overall students whose first names appear earlier or later in the Vietnamese alphabet are equally likely to attend an elite school (Figure 3). Likewise, no information about NHSE performance can be gleaned from one's first name: in all eight subjects, test scores are distributed around the sample mean across the name range (Figure 4). Appendix Figures B2 and B3 repeat this analysis for last names, yielding similar evidence.

Table 4 Summary Statistics on Test Room Characteristics

A Pre-reform

		Physics	Chemistry	Biology	Foreign Language	Math	Vietnamese	Geography	History
No. HS with students in room	*N*	7,577	6,888	7,290	12,549	12,551	12,552	8,534	7,368
	Mean	4.47	3.16	3.59	3.60	3.60	3.60	2.69	4.19
	SD	2.24	2.13	2.13	2.17	2.17	2.17	1.30	2.44
	Median	4.00	2.00	3.00	3.00	3.00	3.00	3.00	4.00
No. students in room	*N*	7,577	6,888	7,290	12,549	12,551	12,552	8,534	7,368
	Mean	23.78	23.75	23.75	23.76	23.68	23.71	23.70	23.68
	SD	1.47	1.64	1.67	1.60	1.64	1.64	1.87	1.48
	Median	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00
$\mathbf{1}\{\text{Room has elite(s)}\}$	*N*	7,577	6,888	7,290	12,549	12,551	12,552	8,534	7,368
	Mean	0.85	0.64	0.72	0.72	0.72	0.72	0.60	0.79
	SD	0.35	0.48	0.45	0.45	0.45	0.45	0.49	0.40
	Median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. elites in room	*N*	7,577	6,888	7,290	12,549	12,551	12,552	8,534	7,368
	Mean	6.41	6.49	6.50	6.46	6.44	6.45	6.50	6.36
	SD	5.34	7.30	6.61	6.58	6.57	6.57	7.29	6.08
	Median	6.00	4.00	5.00	5.00	5.00	5.00	5.00	5.00
Frac. elites in room	*N*	7,577	6,888	7,290	12,549	12,551	12,552	8,534	7,368
	Mean	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
	SD	0.23	0.31	0.28	0.28	0.28	0.28	0.31	0.26
	Median	0.26	0.22	0.22	0.22	0.22	0.22	0.22	0.22
Avg. score elites in room	*N*	6,475	4,440	5,242	9,046	9,047	9,048	5,139	5,847
	Mean	6.40	7.81	6.98	5.27	7.13	5.63	6.20	5.23
	SD	1.78	1.65	1.20	1.68	1.98	1.06	0.79	1.68
	Median	6.28	8.25	7.00	5.00	7.62	5.75	6.25	5.39

B Post-reform

		Physics	Chemistry	Biology	Foreign Language	Math	Vietnamese	Geography	History
No. HS with students in room	*N*	908	908	908	2,488	2,669	2,637	1,909	1,909
	Mean	2.43	2.47	2.43	2.68	2.92	2.85	2.86	2.90
	SD	0.76	0.83	0.80	0.79	1.02	0.92	0.95	0.99
	Median	2.00	2.00	2.00	3.00	3.00	3.00	3.00	3.00
No. students in room	*N*	908	908	908	2,488	2,669	2,637	1,909	1,909
	Mean	22.46	22.62	22.37	23.50	23.50	23.44	23.11	23.20
	SD	2.98	2.84	2.98	1.84	1.81	1.90	2.41	2.30
	Median	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00
$\mathbf{1}\{\text{Room has elite(s)}\}$	*N*	908	908	908	2,488	2,669	2,637	1,909	1,909
	Mean	0.62	0.63	0.63	0.62	0.63	0.62	0.62	0.62
	SD	0.48	0.48	0.48	0.49	0.48	0.49	0.48	0.48
	Median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. elites in room	*N*	908	908	908	2,488	2,669	2,637	1,909	1,909
	Mean	8.15	8.21	8.14	7.38	7.09	7.04	6.43	6.44
	SD	7.87	7.90	7.84	7.22	7.03	7.07	6.78	6.77
	Median	8.00	8.00	8.00	7.00	7.00	6.00	5.00	5.00
Frac. elites in room	*N*	908	908	908	2,488	2,669	2,637	1,909	1,909
	Mean	0.37	0.37	0.38	0.33	0.31	0.31	0.29	0.29
	SD	0.35	0.35	0.35	0.32	0.31	0.31	0.30	0.30
	Median	0.39	0.39	0.39	0.30	0.30	0.30	0.26	0.26
Avg. score elites in room	*N*	566	571	570	1,543	1,672	1,633	1,189	1,190
	Mean	5.54	5.43	4.46	4.10	5.47	6.01	5.93	4.16
	SD	0.92	0.95	0.73	1.28	1.32	0.94	0.72	0.82
	Median	5.63	5.44	4.45	3.82	5.51	6.04	5.94	4.02

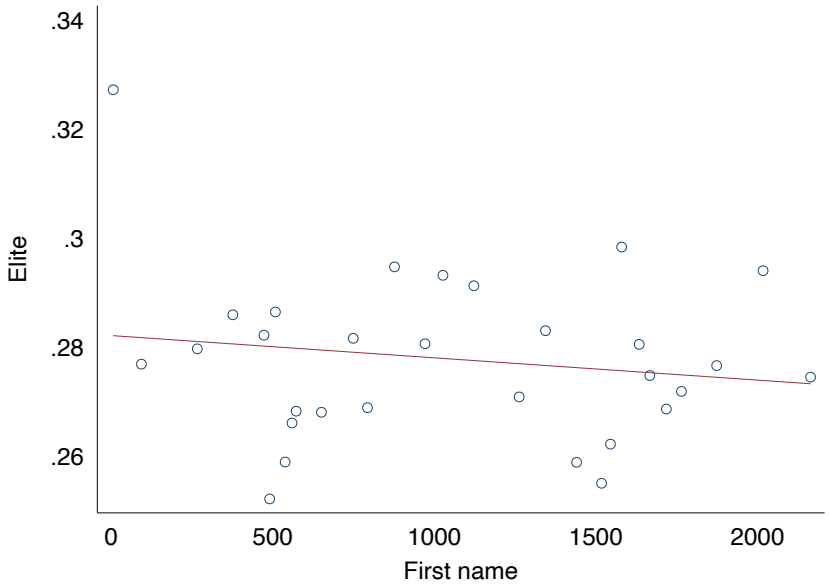
Notes: This table shows the student composition at the test room level. Panel A and Panel B summarize the pre-reform (2007-2013) and post-reform (2018-2019) period, respectively. For each period, the underlying sample consists of students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). All test scores are on a 0-10 grading scale.

TABLE 5 Correlations between Student Names and Scholastic Aptitude

	Elite	Physics	Chemistry	Biology	Foreign Language	Math	Vietnamese	Geography	History
First name	−0.039	−0.023	−0.006	−0.021	−0.068	−0.051	−0.005	−0.034	−0.008
Last name	0.007	−0.012	0.017	0.015	0.027	0.016	0.036	0.038	0.018

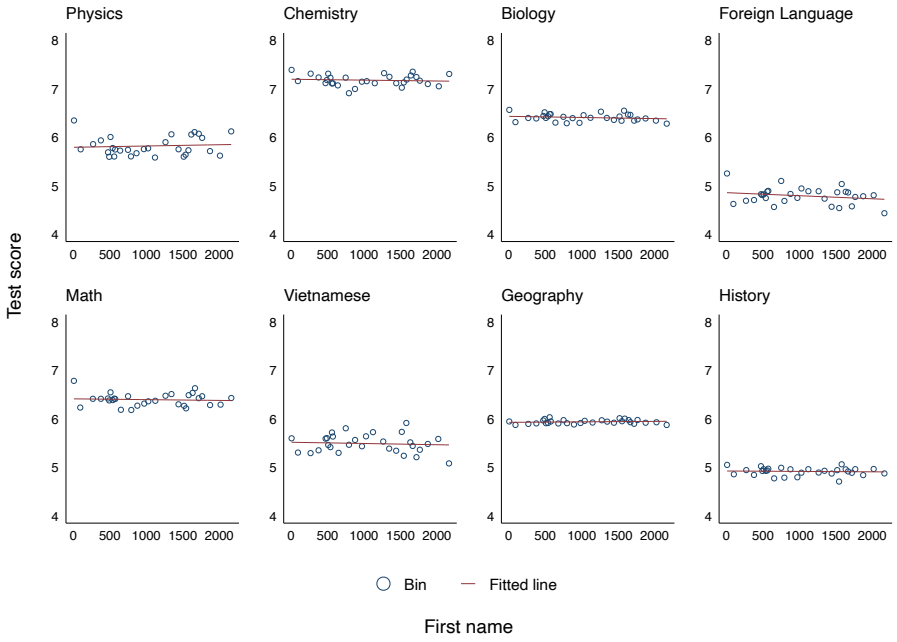
Notes: This table reports the pairwise correlations between student names and academic performance, measured by attendance at an elite high school and NHSE test scores. The underlying sample includes all students who took the NHSE between 2002 and 2019.

FIGURE 3 Binned Scatterplot of Elite Status by First Name



Notes: This figure shows the binned scatterplot of elite status among students by first name. The underlying sample includes all students who took the NHSE between 2002 and 2019.

FIGURE 4 Binned Scatterplots of Test Scores by First Name



Notes: This figure shows the binned scatterplots of test scores among students by first name. The underlying sample includes all students who took the NHSE between 2002 and 2019.

4 Effects of Density of Elite Students

This section investigates the impacts of the quantity of elite students in a test room on the test performance of non-elite peers across different subjects and the evolution of these effects following the 2015 exam redesign. I split non-elite students in the province into two subsamples, depending on when they took the NHSE: 2007-2013 (pre-reform) or 2018-2019 (post-reform).

4.1 Econometric Specifications

For each period and subject, the main specification is a pooled cross-sectional regression:

$$Score_{ij} = EliteQuantity_{ij} + \gamma \boldsymbol{X}_i' + \delta_{sy} + \xi_{cy} + \epsilon_{ij}, \tag{1}$$

where i denotes a non-elite student, j a test subject, s the student's high school, c their test center, and y year. Elite quantity is measured by either the presence or the relative size of elite students in a test room: $ElitesInRoom_{ij}$ is a dummy equal to 1 if student i's room has any elite students and 0 otherwise whereas $FracElitesInRoom_{ij}$ gives the share of elite

students among student i's roommates. For "cheatable" subjects,[7] the marginal effects of elite quantity on non-elite students are expected to be positive pre-reform and null or possibly negative post-reform while for the other subjects, they are hypothesized to be non-positive in both periods. The vector of covariates $\boldsymbol{X}_i$ contains demographic observables (indicators for gender, ethnic minority, and socioeconomic status that determines priority and any bonus points) and academic standing in grade 12 (high- or low-achieving). To control for time-varying factors that may affect different schools and test centers differently, I include both *school* × *year* fixed effects (δ_{sy}) and *test center* × *year* fixed effects (ξ_{cy}). Lastly, ϵ_{ij} is the error term. Test scores are standardized by subject within each period and standard errors are clustered at the *test room* × *year* level.

4.2 Results

4.2.1 Average Effects of Elite Quantity on Test Scores

Table 6 presents the estimates of the effects of elite presence on the test performance of non-elite peers. Before the reform (Panel A), non-elite students sharing a room with at least one elite test-taker scored, on average, 0.23 sd higher in Chemistry, 0.14 sd higher in Physics, 0.11 sd higher in Math, 0.10 sd higher in Biology, and 0.07 sd higher in History. However, they scored 0.15 sd *lower* in Geography. No significant effects were observed in Foreign Language or Vietnamese. After the reform (Panel B), most of these effects disappeared, with the exception of Math and Physics. The presence of elite peers continued to boost non-elite students' Math scores by 0.10 sd, but *reduced* their Physics scores by 0.07 sd.

Looking beyond the extensive margin, Table 7 reports the average effects of elite density. Panel A shows that, in the pre-reform period, a 10 pp increase in the share of elite students in a room led to an average increase in test scores of 0.04 sd in Chemistry and 0.03 sd in Physics, Biology, and Foreign Language, as well as 0.02 sd in Math. In contrast, there was a marginally significant *decline* of 0.01 sd in Geography and no detectable effects in Vietnamese or History. Panel B indicates that, in the post-reform period, test score increases were observed only in Foreign Language and Math, amounting to 0.02 sd and 0.03 sd, respectively, for a 10 pp higher fraction of elite roommates. The density of elite students in a room no longer influenced the scores of non-elite students in Physics, Chemistry, or Biology, all multiple-choice tests. Finally, as in the pre-reform period, the write-in tests—Vietnamese, History, and Geography—remained unaffected.

7. I loosely label multiple-choice/quantitative subject tests as "cheatable" because exams of these types are arguably more amenable to cheating than essay and qualitative exams.

Table 6 Average Effects of Elite Presence on Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
$\mathbf{1}\{\text{Room has elite(s)}\}$	0.136^{***}	0.225^{***}	0.096^{**}	-0.052	0.112^{***}	-0.005	-0.148^{***}	0.071^{**}
	(0.032)	(0.072)	(0.043)	(0.042)	(0.029)	(0.023)	(0.047)	(0.030)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.436	0.419	0.274	0.418	0.549	0.424	0.272	0.521
Observations	129,111	117,079	123,567	213,265	212,641	212,897	144,514	124,970

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
$\mathbf{1}\{\text{Room has elite(s)}\}$	-0.071^{*}	-0.013	0.006	-0.058	0.101^{**}	-0.017	0.095	0.029
	(0.038)	(0.033)	(0.068)	(0.065)	(0.050)	(0.055)	(0.072)	(0.047)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.235	0.226	0.105	0.251	0.407	0.399	0.202	0.163
Observations	12,660	12,660	12,660	39,825	42,355	42,355	31,458	31,458

Notes: This table presents the results of estimating Equation 1 where the main explanatory variable is the presence of elite students in a non-elite student's test room. Panel A reports the estimates for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). The sample consists of non-elite students who took the NHSE in the indicated years and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7 Average Effects of Elite Density on Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of elites in room	0.335***	0.379***	0.287***	0.263***	0.223***	−0.011	−0.083*	−0.011
	(0.042)	(0.061)	(0.054)	(0.048)	(0.033)	(0.030)	(0.049)	(0.042)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.437	0.420	0.275	0.418	0.549	0.424	0.272	0.521
Observations	129,111	117,079	123,567	213,265	212,641	212,897	144,514	124,970

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of elites in room	−0.075	−0.006	−0.095	0.169***	0.274***	−0.023	0.053	0.037
	(0.092)	(0.101)	(0.115)	(0.058)	(0.059)	(0.059)	(0.073)	(0.066)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.235	0.226	0.105	0.251	0.408	0.399	0.202	0.163
Observations	12,660	12,660	12,660	39,825	42,355	42,355	31,458	31,458

Notes: This table presents the results of estimating Equation 1 where the main explanatory variable is the density of elite students in a non-elite student's test room. Panel A reports the estimates for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). The sample consists of non-elite students who took the NHSE in the indicated years and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Heterogeneity I next explore possible heterogeneous effects by student characteristics.

By scholastic aptitude Based on their overall GPA in grade 12, students are classified as either high- or low-achieving. Using this classification, I divide the sample of non-elite students by academic standing and estimate Equation 1 on the resulting subsamples, again separately by period. The results are summarized in Appendix Tables C3 and C4. It is evident that prior to the reform, both high-achieving and low-achieving non-elite students benefited, to generally similar extents, from access to elite peers. After the reform, the positive effect of elite existence found in Math earlier was driven by low-achieving students (Appendix Table C3, cf. Table 6), while the positive impacts of elite density on Foreign Language and Math were due to high-achieving students (Appendix Table C4, cf. Table 7).

By gender Similar heterogeneity results by gender can be found in Appendix Tables C5 and C6. Before the reform, female students gained more from the presence of elite peers than male students. In addition, female (male) students were responsible for the overall positive effect on Foreign Language (Math) observed after the reform, as previously reported in Table 6. There were no gender differences otherwise.

4.2.2 Quantile Effects of Elite Density on Test Scores

The impacts of elite density may not be constant across its distribution. To account for this possibility and to assess potentially nonlinear effects, I rewrite Equation 1 as:

$$Score_{ij} = f\left(FracElitesInRoom_{ij}\right) + \gamma \boldsymbol{X}_i' + \delta_{sy} + \xi_{cy} + \epsilon_{ij} \tag{2}$$

and use *binscatter/binsreg*, the nonparametric Binscatter Least Squares Estimation procedure developed by Cattaneo et al. 2024, to estimate this equation. Throughout, binning is set to be IMSE-optimal and is specific to the underlying estimation sample. Standard errors are clustered at the *test room* × *year* level.

Figure 5 illustrates the estimation results. Each panel shows a binned scatterplot of the effects of the share of same-room elite peers on the test scores of non-elite students in a given subject. For most subjects, the optimal, data-driven number of bins is three. There is little evidence of non-constant effects across the distribution of elite density, as the binned coefficients in each subplot are not statistically different from each other. Notwithstanding, prior to the reform, the larger the group of elite students in the room, the greater the magnitude of elite peer effects. According to Panel A, for example, a 10 pp increase in elite density raised the Physic scores of non-elite students by 0.03 sd in a room where 14% of the other students were elite. This gain grew to 0.04 sd and 0.07 sd in rooms with 43% and 76%

elite peers. Similar patterns are observed in Biology (Panel C) and Foreign Language (Panel D) whereas the score gains in Chemistry (Panel B) from a 10 pp increase in the share of elite students were relatively uniform, amounting to about 0.04 sd across the distribution.

Panels E through H of Figure 5 report the results for the remaining subjects. I first discuss the effects on Math. Regardless of the test format (write-in pre-reform or multiple-choice post-reform), the Math exam primarily tests problem-solving skills. Hence, Math questions in the NHSE are highly standardized, leaving little room for individualization of solutions. This makes cheating both easier (lower costs of cheating in real time) and more profitable (lower risks of being discovered afterward when exams are graded). Thus, cheating on Math seems as likely as cheating on multiple-choice tests. Indeed, Panel E shows that prior to the reform, the effects on Math followed a pattern similar to that of the effects on the multiple-choice subjects: being surrounded by more elite peers raised a non-elite student's Math score. In contrast, no such effects were found in the humanities: Vietnamese, Geography, and History. In a few cases (Vietnamese and History in elite-dense test rooms), the estimates are *negative*, though insignificant. Overall, however, none of the estimates in Panels F, G, and H are statistically different from zero: the null impact prevails regardless of how many elite students were in the test room.

Interestingly, after the exam redesign in 2015, the positive effects of elite density on the multiple-choice subject tests all but disappeared. Panel A of Figure 5 shows that lower densities of same-room elite students did not affect non-elite students' Physics scores. Yet as the fraction of elite roommates approached unity, the non-elite students fared *worse* in Physics. Next, panels B and C demonstrate that neither non-elite students' Chemistry scores nor their Biology scores were affected by the number of elite peers present in the test room. By comparison, Panel D suggests that the impacts on Foreign Language followed an inverted U-shape: null at low and high densities of elite peers, but positive at medium densities.

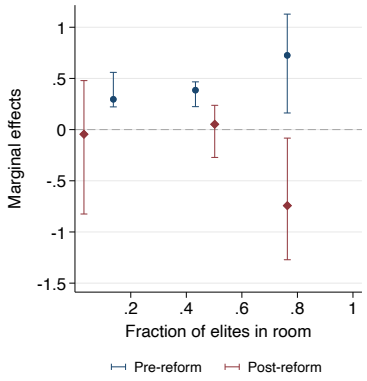
It is clear from the remaining panels that the lack of effects on the humanities subjects carried over to the post-reform period. Finally, Panel E on one hand indicates that after the 2015 exam revamp, elite students continued to have a positive impact on non-elite peers' achievement in Math. On the other hand, such an effect emerged only when there were a high number of elite students in the test room.

Heterogeneity To investigate heterogeneous effects, I estimate Equation 2 again on the subsamples divided by baseline academic ability or gender.

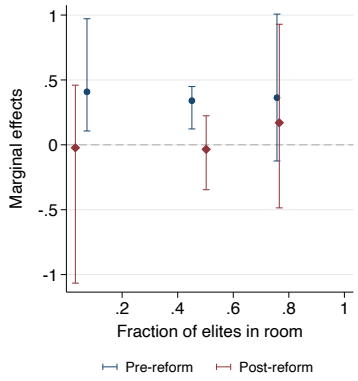
By scholastic aptitude Appendix Figure B4 plots the results by academic ability. For the multiple-choice and physical science test subjects, there exists significant heterogeneity in

FIGURE 5 Quantile Effects of Elite Density on Test Scores

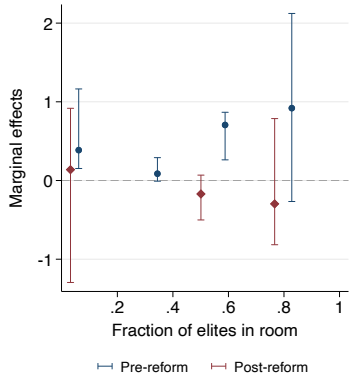
(A) Physics



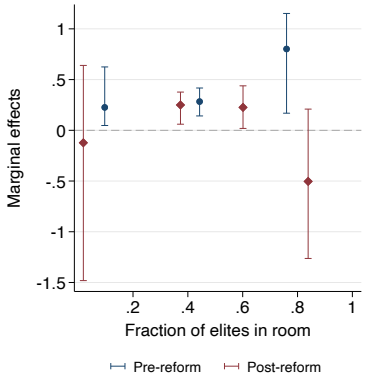
(B) Chemistry



(C) Biology



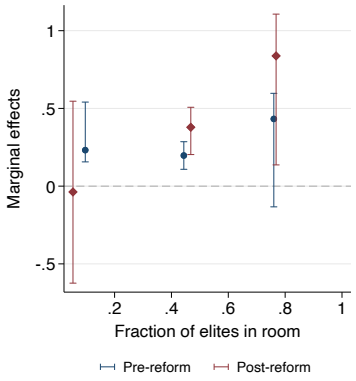
(D) Foreign Language



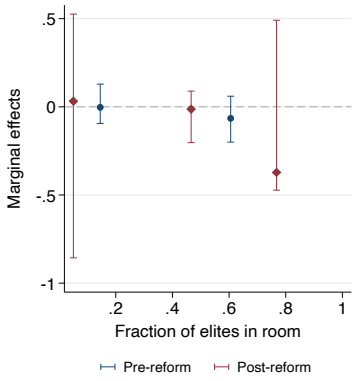
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2. The number of bins is chosen separately by period and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE 5 Quantile Effects of Elite Density on Test Scores (Continued)

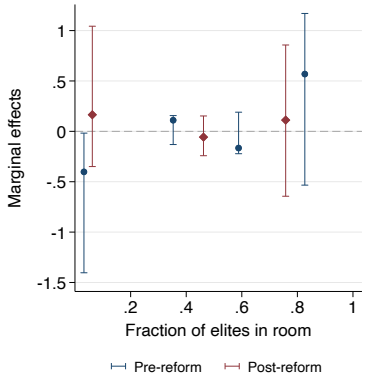
(E) Math



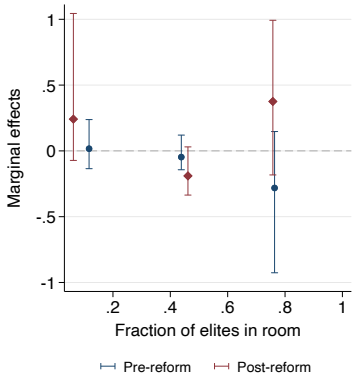
(F) Vietnamese



(G) Geography



(H) History



Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2. The number of bins is chosen separately by period and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

elite peer effects by baseline ability. Pre-reform, both groups of non-elite students benefited from sharing the same test room with elite peers, but low-achieving non-elite students gained more than high-achieving schoolmates. Post-reform, however, access to more elite peers no longer improved test performance, especially for students with lower baseline ability. Lastly, juxtaposing Appendix Figure B4 with Figure 5 reveals that the few positive effects in Foreign Language (Panel D) and Math (Panel E) documented therein were driven by high-achieving non-elite students.

By gender The share of elite students in the test room did not appear to affect the test achievement of non-elite students differently by gender. As shown in Appendix Figure B5, largely similar gains from access to elite peers accrued to male and female non-elite students, both pre- and post-reform. Across the eight subjects, the patterns of effects for male and female students closely mirror each other.

Robustness The estimated effects of elite density are robust to a range of sensitivity checks. For example, dropping covariates from Equation 1 makes no qualitative differences (Appendix Figure B6). Other checks, such as redefining elite status (see Section 2) and estimating separate regressions by year instead of pooling the data, further corroborate the robustness of the main results. Details on these additional tests are available upon request.

5 Effects of Quality of Elite Students

Conditional on sharing the test room with elite students, did non-elite students benefit more from high-achieving than low-achieving elite peers? Test achievement might not be homogeneous among elite test-takers: different students may be good at different subjects. Moreover, despite school prestige, it is possible that an elite student performs worse than a non-elite peer in a particular subject test. As such, the quality of elite peers is potentially important.

To evaluate this hypothesis, I first visualize the relationship between the test scores of non-elite students and the average test scores of elite peers in the same room in a series of binned scatterplots. Before binning and plotting, I residualize both test scores on the share of elite students in a room, non-elite students' characteristics as included in the vector of covariates $\boldsymbol{X}$, and the fixed effects in Equation 1. Figure B7 organizes the resulting graphs by period: pre-reform (Panel A) and post-reform (Panel B). Both panels suggest a linear relationship between the test scores of the two groups of students. However, the direction of this relationship changed between the two periods of interest. Prior to the reform, across all subjects, the higher the scores of elite students, the higher those of their non-elite peers.

In contrast, except for Math and Vietnamese, the scores of the two groups were no longer correlated after the reform.

5.1 Econometric Specifications

Ordinary Least Squares Motivated by the graphical evidence in Figure B7, I estimate the following linear specification:

$$(3)\quad Score_{ij} = \dot{\alpha} + \dot{\beta}\, AvgScoreElitesInRoom_{ij} + \dot{\omega}\, FracElitesInRoom_{ij} + \dot{\gamma}\, \boldsymbol{X}_i' + \dot{\delta}_{sy} + \dot{\xi}_{cy} + \dot{\epsilon}_{ij}.$$

The variables common to Equation 1 retain their definitions. The new regressor is the average score in subject j of all elite students in non-elite student i's test room, $AvgScoreElitesInRoom_{ij}$. Of primary interest is the coefficient on this variable, $\dot{\beta}$. In the absence of endogeneity, for cheatable subjects, $\dot{\beta}$ is expected to be positive pre-reform—all else equal, for a non-elite cheater, a high-performing elite student would be a more useful resource than a low-performing elite student—and non-positive post-reform.

A caveat of Equation 3 is that $AvgScoreElitesInRoom_{ij}$ is likely endogenous. One obvious source of endogeneity is the reflection problem (Manski 1993). While an elite student might raise the test performance of a non-elite student in the same room, in reverse, the non-elite student might also affect the elite student's test outcome. The direction of this non-elite peer effect is unclear. On the one hand, being surrounded by lower-performing peers might boost the elite student's confidence, resulting in a higher test score. On the other hand, if the non-elite student was seeking to cheat by copying the elite student's answers, then the non-elite peer could become a nuisance to the elite student, impairing their test performance. In general, endogeneity due to reflection is notoriously difficult to overcome (Moffitt 2000, Epple and Romano 2011, Sacerdote 2014).

As a result, the coefficient on the average test score of elite students in the same test room, $\dot{\beta}$, may not admit a causal interpretation. One obvious way to alleviate the above endogeneity concerns is to use student ability measured before the exam in question to proxy for student quality. Unfortunately, my data lacks a subject-specific measure of baseline scholastic aptitude: I only observe a categorical variable of academic ranking in grade 12 (up to four distinct values). In addition to being coarse, this academic ranking is school-specific, hence, not immediately comparable across schools.

To make headway in quantifying the contemporaneous effects of elite students' quality on non-elite students' test performance, I now turn to a subset of the sample. This subsample differs from the full sample in two important ways. First, it only retains the students who

went to high school in the capital city.[8] Second, the pre-reform component is two years shorter: 2007-2011 (as opposed to 2007-2013). I drop 2012-2013 because the two specialized high schools happened to constitute one same test committee in those years. There was thus no mixing between specialized and non-specialized students in these years.

Instrumental Variables The province in question has two specialized high schools, both located in its capital city. As described in Section 2, admission to specialized high schools is highly competitive. Like their counterparts elsewhere in the country, the two specialized high schools in this province each offer a variety of majors. One of the schools has a larger course menu: students can specialize in subjects in physical science, social science, or humanities. The other school traditionally focused on Math and Physics. Chemistry, Biology, Foreign Language, and Vietnamese were not offered as subjects of specialization until recently.

At either school, students are organized into classrooms based on their majors. Importantly, a specialized student's classroom name indicates which subject they specialize in. Exploiting this correspondence between the classroom name and the subject of specialization, I estimate the following IV model on the subsample of students in the capital city:

$$AvgScoreElitesInRoom_{ij} = a^1 + b^1\, FracSubjectElitesInRoom_{ij} + c^1\, FracElitesInRoom_{ij} + g^1\, \boldsymbol{X}_i' + \delta^1_{sy} + \xi^1_{cy} + e^1_{ij}, \tag{4}$$

$$Score_{ij} = a^2 + b^2\, \widehat{AvgScoreElitesInRoom_{ij}} + c^2\, FracElitesInRoom_{ij} + g^2\, \boldsymbol{X}_i' + \delta^2_{sy} + \xi^2_{cy} + e^2_{ij}. \tag{5}$$

Equations 4 and 5 specify the first and second stage, respectively. The numerical superscripts on the parameters and error terms refer to the relevant stage. Most variables and parameters are retained from Equations 1 and 3. The new addition is the instrument for the average subject score of same-room elite students: $FracSubjectElitesInRoom_{ij}$, or the share of subject-specific elite students present in a non-elite student's room. Like $FracElitesInRoom_{ij}$, this instrument is exogenous to non-elite test scores on the grounds of quasi-random assignment of students into test rooms within a test center. In addition, since classroom information was never published in examinee rosters, it is plausible that the exclusion restriction is satisfied. Finally, within a a test room, the fraction of elite students majoring in a particular subject is expected to predict the average test score of all elite students in that subject.

8. Appendix Tables C1 and C2 summarize student and test room characteristics, respectively, for this subsample. By and large, it is comparable to the full sample of the entire province.

In the second stage, the predicted average elite performance is used to explain the performance of non-elite peers, conditional on the general density of elite students and all the other usual covariates. The parameter of interest is b^2. Like its counterpart $\dot{\beta}$ in Equation 3, b^2 is hypothesized to be positive. Unlike $\dot{\beta}$, however, b^2 can be interpreted as causal.

5.2 Results

OLS Estimation Table 8 tabulates the results of estimating Equation 3 separately by period. As expected, holding constant same-room elite density and individual observable characteristics, the test scores of non-elite students were positively correlated with the average test scores of elite roommates. Panel A shows that before the reform, a one sd higher in the average test score of elite students in Physics, Chemistry, Biology, Math, Geography, and History was associated with an approximately 0.2 sd increase in the test score of non-elite peers in the same subjects. The score boost was largest in Foreign Language at 0.5 sd and smallest in Vietnamese at 0.1 sd. Panel B, however, indicates that following the reform, the conditional correlations between elite and non-elite test scores lost statistical significance and weakened toward zero.

IV Estimation Are the correlations in test scores between the two groups of students documented above completely fraught with endogeneity? To obtain causality, I use the subsample of students in the capital city to estimate the IV model characterized by Equations 4-5, separately by period.[9] Table 9 reports the first stage results.[10] Prior to the reform (Panel A), the share of subject-specific elite students in a room strongly predicted the test performance of all elite students in the same room in all but two subjects (Chemistry and Geography). All else equal, a 10 pp increase in the proportion of subject elite students in a room raised that room's average elite test score by 1.4 sd in Biology, 1.2 sd in Foreign Language, 0.9 sd in Vietnamese, 0.5 sd in History, 0.4 sd in Physics, and 0.1 sd in Math. This predictive power remained strong after the reform (Panel B): the same increase in the share of subject-specific elite students in a room corresponded to a 0.4-1 sd increase in the average test score of all elite students in that room.

Table 10 reports the second stage results, using the density of subject elite students to instrument for the quality of all elite students in a room. Panel A shows that higher-achieving elite students boosted non-elite peers' performance in all multiple-choice subjects except

9. Prior to the reform, the test scores of the two groups of students in this subsample were also positively correlated. The magnitude of correlation was largely similar to that found for the full sample covering the entire province. After the reform, however, such correlations evaporated. See Appendix Table C7 for details.

10. The reduced form estimates are collected in Appendix Table C8.

Table 8 Relationship between Elite Quality and Non-elite Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	0.208***	0.254***	0.250***	0.472***	0.170***	0.130***	0.252***	0.239***
	(0.007)	(0.009)	(0.008)	(0.009)	(0.006)	(0.005)	(0.010)	(0.008)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.447	0.440	0.287	0.506	0.543	0.418	0.295	0.499
Observations	103,517	59,652	75,568	131,314	130,710	130,949	65,075	89,580

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	0.016	0.008	0.039**	0.030*	0.132***	0.082***	0.022*	−0.004
	(0.015)	(0.018)	(0.019)	(0.017)	(0.017)	(0.011)	(0.011)	(0.012)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.273	0.229	0.110	0.342	0.442	0.376	0.181	0.119
Observations	5,208	5,261	5,240	17,670	19,400	19,162	14,960	14,979

Notes: This table presents the relationship between the quality of same-room elite peers and non-elite students' test scores. Panel A reports the regression results for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). All columns control for the density of same-room elite peers. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE 9 **First Stage:** Subject of Specialization and Overall Performance among Elites

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of subject elites in room	3.996***	3.248	13.504***	12.054***	1.465**	8.508***	1.336	5.176**
	(0.719)	(2.268)	(1.809)	(1.336)	(0.623)	(1.249)	(1.728)	(2.109)
Frac. of elites in room	0.525	1.222**	0.584	0.066	0.404	−0.606*	−0.248	0.235
	(0.372)	(0.619)	(0.464)	(0.306)	(0.379)	(0.310)	(0.390)	(0.411)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	30.860	2.050	55.706	81.382	5.532	46.429	0.597	6.024
Adj. R^2	0.302	0.134	0.187	0.396	0.207	0.372	0.341	0.445
Observations	13,046	7,470	8,959	16,431	16,374	16,403	8,408	11,353

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of subject elites in room	5.275***	5.100***	6.050***	3.923***	3.831***	9.976***	8.568***	8.180**
	(1.210)	(1.182)	(1.361)	(0.667)	(0.693)	(1.508)	(2.117)	(3.428)
Frac. of elites in room	0.004	−0.258	−1.463**	−1.980***	−0.723	−0.207	0.762	0.692
	(0.888)	(0.773)	(0.708)	(0.409)	(0.522)	(0.493)	(0.835)	(0.826)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	19.010	18.629	19.758	34.629	30.593	43.767	16.386	5.694
Adj. R^2	0.439	0.324	0.442	0.666	0.606	0.503	0.271	0.375
Observations	1,572	1,577	1,577	3,230	3,514	3,444	2,312	2,312

Notes: This table presents the first stage results of estimating the IV model characterized by Equations 4-5, using the share of subject-specific elites in a test room as an instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

for Chemistry. Holding fixed the fraction of elite students in a room, a one sd increase in their Physics, Biology, or Foreign Language scores raised non-elite peers' scores in the same subjects by 0.4-0.5 sd on average. Intriguingly, as revealed in Panel B, these positive effects died out after the exam revamp in 2015. The coefficient on the average test score of elite students is only significant for Physics and Vietnamese. In the case of Physics, however, the sign is negative: access to higher-performing elite peers *lowered* non-elite students' Physics test score. In contrast, elite equality had a positive impact on non-elite students' Vietnamese test score—an effect that was absent in the pre-reform period.

TABLE 10 **Second Stage:** Effects of Elite Quality on Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	0.451***	−1.015	0.446***	0.492***	−0.080	−0.054	0.205	0.004
	(0.102)	(0.943)	(0.071)	(0.055)	(0.174)	(0.057)	(0.595)	(0.198)
Frac. of elites in room	−0.043	1.279	0.193	0.596***	0.115	−0.044	−0.192	0.063
	(0.228)	(1.398)	(0.253)	(0.165)	(0.192)	(0.107)	(0.224)	(0.213)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.040	−1.482	0.137	0.234	0.096	0.166	0.109	0.045
Observations	13,046	7,470	8,959	16,431	16,374	16,403	8,408	11,353

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	−0.139*	−0.016	−0.028	0.063	0.107	0.157**	−0.015	0.079
	(0.073)	(0.109)	(0.094)	(0.057)	(0.065)	(0.072)	(0.091)	(0.140)
Frac. of elites in room	0.269	−0.070	−0.058	0.174	−0.066	0.102	0.122	0.378
	(0.303)	(0.298)	(0.339)	(0.136)	(0.133)	(0.168)	(0.257)	(0.312)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.018	−0.010	0.004	0.009	−0.000	0.176	0.002	−0.002
Observations	1,572	1,577	1,577	3,230	3,514	3,444	2,312	2,312

Notes: This table presents the second stage results of estimating the IV model characterized by Equations 4-5, using the share of subject-specific elites in a test room as an instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Heterogeneity The lack of variation in baseline academic ability among non-elite test-takers in the capital city precludes subsample analysis along this dimension. I therefore focus on subsample analysis by gender. Panel A of Table C10 reveals few gender differences in the pre-reform period: both male and female non-elite students benefited, to similar extents, from

elite peers of high quality. Post-reform, however, gender gaps emerged in Physics, Math, and Vietnamese (Panel B, ibid). In particular, the negative effect previously found in Physics (Column 1, Panel B of Table 10) was due to male students whereas female students drove the positive effect found in Vietnamese (Column 6, ibid).

Robustness The above estimates of peer effects are robust to the choice of instrument. An alternative instrument for the average test performance of elite students is the share of elite students within a field of study (as opposed to a narrow subject). As described in Section 2, Vietnam's college admission exams are organized by blocks of subjects within either physical science (Math, Physics, Chemistry, and Biology) or social science and humanities (Vietnamese, History, Geography, and Foreign Language). Since most specialized students are college aspirants, those who major in a certain subject likely perform well in the other subjects of the same field. The implication is that non-elite students who share the same test room with, for example, Math majors can score higher not only in Math, but also in Physics, Chemistry, or Biology.

I proceed by matching specialized students to their fields of study and calculate for each test room the leave-one-out fraction of elite test-takers specialized in a given field. I then use this share as a new instrument to re-estimate the IV model of elite peer quality (Equations 4-5). Appendix Table C11 reports the first stage results. Pre-reform, the alternative instrument predicted elite students' test performance across all subjects except Geography and History, although not as strongly as the main instrument (cf. Table 9). However, post-reform, it became weaker in Physics, Chemistry, and Biology. Nevertheless, the second stage results, reported in Appendix Table C12, corroborate the qualitative conclusion from Table 10: elite peer quality mattered, but only at a particular time (pre-reform) and in a particular set of subject tests (multiple-choice). Lastly, heterogeneity analysis by gender in Appendix Table C14 indicates that both male and female non-elite test-takers benefited from higher quality of elite peers, with few gender differences. However, the score gains occurred in different subjects depending on the period: Physics, Chemistry, Biology, and Foreign Language pre-reform, versus Foreign Language, Math, and Vietnamese—namely the three required subjects for high school graduation—post-reform.

6 Discussion and Conclusion

Taken together, the results in Sections 4 and 5 reinforce each other, highlighting two curious patterns. First, positive peer effects were present prior to the reform, but largely disappeared afterward. Second, the positive peer effects found before 2015 were confined to

the multiple-choice or quantitative test subjects. This paper contends that these patterns suggest mass cheating between elite and non-elite students when the exams were low-stakes. It is worth emphasizing that such cheating is not merely conceptual, but feasible in practice. As illustrated in Figure 1, the test room had an open layout without physical partitions between test-takers. The lack of physical barriers made it easy for students in the same room to pass answers to and through neighbors. The invigilator might not detect this clandestine communication or choose to ignore it. Notwithstanding, at first thought, cheating might not be the most obvious explanation of the peer effects that this paper has uncovered. Below, I address several competing hypotheses to cheating and argue that none of them is plausible in my setting.

Legitimate Peer Effects Cheating by copying from other test-takers, albeit dishonest, is a channel of peer effects. Theoretically, peer effects can operate through other, legitimate mechanisms. For example, being assigned to the same test room with many elite students may motivate a non-elite student to work harder on the exam. In this case, the mere presence of high-achieving peers compels the non-elite student to make greater efforts. This hypothesis is tempting: with public rosters, any student could observe ahead of time who would be in their test room, opening the possibility of emulation. In my context, however, it is misguided to attribute better test outcomes to a desire to emulate high-achieving peers.

First, it is unclear why with greater ambition, higher test scores were only realized in the quantitative and multiple-choice subject tests. Either no effects or negative effects of elite quantity were found in the open-ended subject tests. Since these essay tests last longer and the grading rubrics are less clear-cut than the multiple-choice exams, one would expect extrinsic motivation to be more influential for them. Moreover, while a student could easily look up which high school a fellow test-taker attended, they might not know which subjects a certain elite student was good at. The list of test-takers by test room posted at each test center did *not* show the name of a student's classroom, only that of their high school. Yet, as shown in Section 5, having subject- or field-specific elite students in the same room gave non-elite students an edge above and beyond what was given by so-called plain elite students.

Second, assume that consecutive cohorts were similar in cognitive skills and motivation to excel on the NHSE. Then if the higher test scores among non-elite test-takers sharing a test room with elite students were indeed due to ambition and emulation, it would be puzzling to find positive impacts in a period, but nil in a subsequent period. However, this break in effects is precisely what I have discovered. Finally, it should be noted that the NHSE lasts but three days. Unlike classmates who spend years in school together, test-takers have limited opportunities to interact with each other during this short exam window. In light of

the lack of prolonged interaction between test-takers, legitimate peer effects seem unlikely to have arisen.

Instead of external motivation, peer effects can take the form of peer disruption. A non-elite student assigned to a test room with more elite students may be less susceptible to disruptive behavior by fellow test-takers because elite students are generally more disciplined. In the current setting, however, this hypothesis is untenable. In general, the students are well-behaved regardless of their elite status. More importantly, as a national event, the NHSE is subject to heavy policing. It is customary for the local police to patrol test centers (in hallways outside test rooms). As such, any behavioral disturbance should be minimal.

Proctoring NHSE proctors are either local teachers or teachers from another province. The assignment of proctors to test centers and test rooms is a top-down decision made by the provincial DOET. Neither the proctors nor the students have prior knowledge of who will guard which test room in which test session. Most importantly, teachers are not allowed to proctor at the test center where students from their own school take the NHSE. Cross-proctoring thus disincentivizes proctors from directly engaging in cheating.

It is still possible that proctors have abetted existing cheating among students, e.g., by turning a blind eye to it. But even if proctors were complicit, their cheating would be furtive at most. Following the aforementioned scandal in 2006, proctors started facing higher risks of detection if they cheated and more severe punishment if they got caught cheating. These institutional details substantiate the argument that proctor cheating was unlikely to overshadow student cheating in both periods in question, although my current analysis cannot rule out proctors' potential involvement altogether.

Other Test Room Characteristics How probable is it that something else about a test room other than its student composition was responsible for the results? The answer is in the negative. As described in Section 2, NHSE test rooms are standard, regular classrooms in typical school buildings. In addition, there is little variation in physical attributes across classrooms within a test center in a given year; in all estimations, I include *test center* × *year* fixed effects. Finally, as shown in Section 3, there was no variation in the size of a test room (in terms of the number of students) in the data; virtually every room reached the maximum capacity permitted. This leaves the student composition of a test room as the most probable cause of the observed peer effects.

Conclusion This paper conducts "forensic economics" (Zitzewitz 2012) to unmask shrewd examination malpractice by students. As a means of coping with grade pressure, student cheating on exams has received less attention from economists than manipulation

by schools and teachers. Jacob and Levitt 2003 and Borcan, Lindahl, and Mitrut 2017 are two studies that investigate test fraud by students. Contributing to this nascent literature,This paper is a detective investigation into possible mass cheating on Vietnam's high school exit exam, using data from a large province. The misconduct in question is cheating between elite students and non-elite students who sit in the same test room on test day. I hypothesize that pressure to pass the exams incentivizes non-elite students to network with elite peers in order to obtain illicit help from them during the exam. Such *ad hoc* "friendship" may be fleeting, but can be advantageous to the non-elite students while it lasts.

To test this hypothesis, I harness the quasi-random assignment of students from schools of varying quality into test rooms. Quasi-randomness comes from the alphabetization of student names for test administration purposes. On a subset of the data, I also leverage information on each elite student's classroom name, which signifies their subject and field of specialization in high school, in an IV framework to identify the contemporaneous impacts of specialized test-takers' test performance on non-specialized peers'. I argue that these effects, when materializing in particular patterns, are symptomatic of cheating.

I find compelling evidence that mass cheating, as hypothesized, has indeed occurred. Pre-reform, positive effects from elite students to non-elite students in the same room were observed in multiple-choice or quantitative subject tests (Physics, Chemistry, Biology, Foreign Language, and Math). On the extensive margin, taking the exams in a room with at least one elite student raised non-elite test scores by 0.1-0.2 sd on average. On the intensive margin, the effects were by and large constant: a 10 pp increase in the share of elite peers in the same test room led to a 0.02-0.04 sd increase in the test scores of non-elite students. As the 2015 reform streamlined testing and raised the exam stakes, however, these positive effects all but disappeared. In contrast, in both periods, impacts were absent in the essay, humanities tests (Vietnamese, History, and Geography).

Similar patterns were uncovered in the IV analysis, which focuses on the effects of elite quality. High-achieving elite peers were more useful, but only prior to the reform: a one sd increase in the average quality of elite peers results in a 0.4-0.5 sd increase in non-elite test scores. Moreover, these effects were confined to the multiple-choice subjects. Finally, there is considerable heterogeneity by student ability but less by gender. In summary, supported by various institutional details, my results are most consistent with cheating. In particular, neither legitimate peer effects (e.g., a desire to emulate high-achieving peers or disruptions caused by misbehaving test-takers) nor potential cheating by proctors appears to be a plausible alternative explanation.

The present analysis is not without flaws. First, a more granular measure of elite students' baseline ability could strengthen it, for example, by serving as an instrument for elite test performance. In want of such a measure, I instead deduce an elite student's relative strength among the different test subjects from the name of their high school classroom. Since this deduction is only feasible for the student population in the capital city, where the specialized high schools are located, the estimation sample size is reduced. Second, this paper identifies cheating as running from elite students (the benefactors) to non-elite students (the beneficiaries). However, since elite students may not be equally proficient in all areas of study, cheating might have also occurred within the elite group. Investigating this behavior presents a promising avenue for future research. Finally, as in Jacob and Levitt 2003, the statistical evidence presented here can only suggest the existence of academic fraud; no specific individuals can be implicated. Despite its lack of legal applicability, my "postmortem" study contributes to bridging the gap in empirical economic research on cheating, both in educational contexts and beyond.

References

Angrist, Joshua, and Victor Lavy. 2009. "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial." *American Economic Review* 99, no. 4 (September): 1384–1414. (Cited on page 3)

Borcan, Oana, Mikael Lindahl, and Andreea Mitrut. 2017. "Fighting Corruption in Education: What Works and Who Benefits?" *American Economic Journal: Economic Policy* 9, no. 1 (February): 180–209. (Cited on pages 1, 4, 7, 37)

Carnoy, Martin, and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." Publisher: American Educational Research Association, *Educational Evaluation and Policy Analysis* 24, no. 4 (December): 305–331. (Cited on page 3)

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2024. "On Binscatter." *American Economic Review* 114, no. 5 (May): 1488–1514. (Cited on page 23)

Cullen, Julie Berry, and Randall Reback. 2006. *Tinkering Toward Accolades: School Gaming Under a Performance Accountability System.* Technical report w12286. National Bureau of Economic Research, June. (Cited on page 3)

Dang, Hai-Anh, Paul Glewwe, Jongwook Lee, and Khoa Vu. 2023. "What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the 2012, 2015, and 2018 PISA Data." *Economics of Education Review* 96 (October): 102434. (Cited on page 1)

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48, no. 2 (June): 424–455. (Cited on page 10)

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2016. "School Accountability, Postsecondary Attainment, and Earnings." *Review of Economics and Statistics* 98, no. 5 (December): 848–862. (Cited on page 3)

Epple, Dennis, and Richard E. Romano. 2011. "Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence." In *Handbook of Social Economics,* edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 1:1053–1163. North-Holland, January. (Cited on page 28)

Figlio, David N. 2006. "Testing, Crime and Punishment." *Journal of Public Economics* 90, no. 4 (May): 837–851. (Cited on page 3)

Figlio, David N., and Lawrence S. Getzler. 2006. "Accountability, Ability and Disability: Gaming the System?" In *Improving School Accountability,* edited by Timothy J. Gronberg and Dennis W. Jansen, 14:35–49. Advances in Applied Microeconomics. Emerald Group Publishing Limited, January. ISBN: 978-1-84950-446-1 978-0-7623-1351-8. (Cited on page 3)

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2, no. 3 (July): 205–227. (Cited on page 3)

Goslin, David A. 1963. *Search for Ability, The: Standardized Testing in Social Perspective.* Russell Sage Foundation. ISBN: 978-0-87154-357-8. (Cited on page 1)

Grissmer, David W., Ann Flanagan, Jennifer H. Kawata, and Stephanie Williamson. 2000. "Improving Student Achievement: What State NAEP Test Scores Tell Us." Publisher: RAND Corporation (January). (Cited on page 3)

H., A. 2014. "Năm Vụ Bê Bối Thi Tốt Nghiệp Từng Chấn Động Dư Luận." *Zing News* (June). (Cited on pages 2, 5)

Hanushek, Eric A., and Margaret E. Raymond. 2003. "Improving Educational Quality: How Best to Evaluate Our Schools?," edited by Yolanda Kodrzycki, 193–224. Education in the 21st Century: Meeting the Challenges of a Changing World. Boston, MA: Federal Reserve Bank of Boston. (Cited on pages 1, 3)

———. 2005. "Does School Aaccountability Lead to Improved Student Performance?" _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.20091, *Journal of Policy Analysis and Management* 24 (2): 297–327. (Cited on page 3)

Jacob, Brian A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89, no. 5 (June): 761–796. (Cited on page 3)

Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Publisher: Oxford Academic, *The Quarterly Journal of Economics* 118, no. 3 (August): 843–877. (Cited on pages 1, 3, 4, 37, 38)

Jacob, Brian, and Jesse Rothstein. 2016. "The Measurement of Student Ability in Modern Assessment Systems." Publisher: American Economic Association, *The Journal of Economic Perspectives* 30 (3): 85–107. (Cited on page 1)

Kane, Thomas J., and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives* 16, no. 4 (December): 91–114. (Cited on pages 12, 13)

Kiều, Trinh. 2012. "Những Vụ Gian Lận Thi Tốt Nghiệp THPT Đình Đám." *VnExpress* (June). (Cited on pages 1, 5, 8, B1)

Linn, Robert L. 1991. *Test Misuse: Why Is It So Prevalent?* Technical report. ERIC Number: ED340778. September. (Cited on page 1)

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." Publisher: Oxford Academic, *The Review of Economic Studies* 60, no. 3 (July): 531–542. (Cited on page 28)

Marginson, Simon. 2011. "The Confucian Model of Higher Education in East Asia and Singapore." In *Higher Education in the Asia-Pacific: Strategic Responses to Globalization,* edited by Simon Marginson, Sarjit Kaur, and Erlenawati Sawir, 53–75. Dordrecht: Springer Netherlands. ISBN: 978-94-007-1500-4. (Cited on page 1)

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2004.00481.x, *Econometrica* 72 (1): 159–217. (Cited on page 10)

Ministry of Education and Training of Vietnam. 2007. *Circular/Decision: Promulgating the Regulation on Upper Secondary Education Graduation Examination.* (Cited on page 6)

Moffitt, Robert. 2000. "Policy Interventions, Low-Level Equilibria And Social Interactions." *See Durlauf & Young 2001* (June). (Cited on page 28)

Moses, Michele S., and Michael J. Nanna. 2007. "The Testing Culture and the Persistence of High Stakes Testing Reforms." Publisher: Purdue University Press, *Education and Culture* 23 (1): 55–72. (Cited on page 1)

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." Publisher: The University of Chicago Press, *Journal of Political Economy* 119, no. 1 (February): 39–77. (Cited on page 3)

Nguyễn, Linh C. 2016. "The 'Achievement Epidemic' in Vietnam Schools." *GlobalVoices* (July). (Cited on page 5)

Quân, Lê, and Đan Hạ. 2012. "Lộ Clip Gian Lận Thi Cử Tại Phòng Thi Khác Của Trường Đồi Ngô." Section: Giáo dục, *Thanh Niên* (June). (Cited on page B1)

Sacerdote, Bruce. 2014. "Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?" _eprint: https://doi.org/10.1146/annurev-economics-071813-104217, *Annual Review of Economics* 6 (1): 253–272. (Cited on page 28)

Stankov, Lazar. 2010. "Unforgiving Confucian culture: A breeding ground for high academic achievement, test anxiety and self-doubt?" *Learning and Individual Differences* 20, no. 6 (December): 555–563. (Cited on page 1)

Thanh, Hà, and Thanh Quốc. 2006. "Gặp 'Giám Thị Tố Cáo Tiêu Cực Thi Cử'." Section: Giáo dục, *Tuổi Trẻ Online* (June). (Cited on page B1)

Zitzewitz, Eric. 2012. "Forensic Economics." Publisher: American Economic Association, *Journal of Economic Literature* 50 (3): 731–769. (Cited on pages 3, 36)

Appendices

A Conceptual Framework

This section develops a stylized game-theoretic model to rationalize the hypothesized cheating behavior among students on test day. I model each test session as a sequential game with two players: a non-elite student (Player 1) and an elite student (Player 2). The non-elite student moves first, choosing whether to cheat by soliciting exam answers from the elite student. In response, the latter decides whether to share solutions with the non-elite peer. Both players are utility maximizers, where utility is additively separable in test scores (t) and net gain from cheating ($g - c$):

$$U_i(t_i, g_i, c_i) = u_i(t_i) + v_i(g_i - c_i)$$

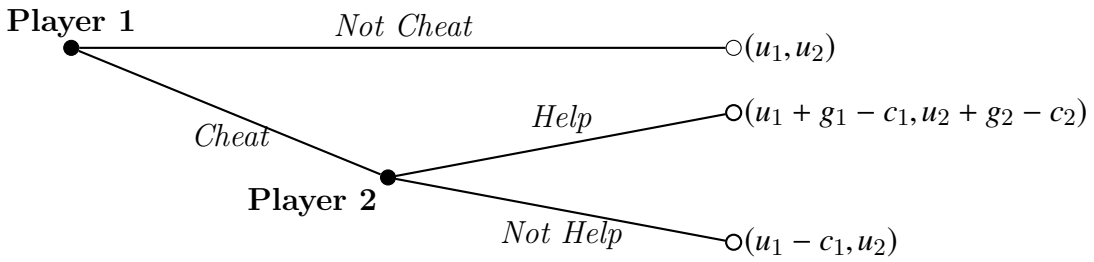
for $i = 1, 2$. The function $u_i(\cdot)$ is increasing in test scores, but differs in form across periods. Prior to the reform, the NHSE was merely a high school exit exam. One needed to pass the exam to take the college admissions exam administered later. Beyond achieving a passing grade, however, NHSE performance is irrelevant to college applications. Therefore, I assume that each player's utility from test scores is a piecewise function:

$$u_i^{\text{pre}}(t_i) = \begin{cases} \underline{u}_i & \text{if } t_i \geq \underline{p}, \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2$, where $\underline{p}$ is the passing grade and $\underline{u}_i$ is a constant. After the reform, student performance on the NHSE is also used for college selection. I hence assume that u_i is now some strictly increasing function $f(\cdot)$ of test scores: $u_i^{\text{post}}(t_i) = f(t_i)$ for $i = 1, 2$.

By cheating or helping the other player cheat, Player i gains g_i at a cost of c_i. For simplicity, I assume that utility from this behavior is linear in both gains and costs: $v_i(g_i, c_i) = g_i - c_i$. The costs of cheating involve the effort required to communicate with the other player and evade proctors' monitoring. These costs vary across subjects: both c_1 and c_2 are lower for multiple-choice and quantitative tests than for essay exams. If Player 1 cheats, he incurs a cost of $c_1 > 0$ but stands to gain $g_1 > 0$ if Player 2 agrees to assist, and zero otherwise. Extending help to Player 1 costs Player 2 an amount of $c_2 > 0$, but generates an altruistic benefit of $g_2 \geq 0$. Without loss of generality, I assume that g_2 was positive prior to the reform, but dropped to zero after the reform raised the stakes of the exam. Figure A1 illustrates the game tree, where superscripts and functional notations are omitted to reduce visual clutter.

FIGURE A1 A Stylized Game of Cheating



The game is solved by backward induction. I first consider the pre-reform period. Suppose Player 1 cheats. If the net gain to Player 2 from helping, $g_2 - c_2$, is positive—which is likely in the multiple-choice tests (where c_2 is low)—Player 2 will choose to help. Anticipating a positive net gain of $g_1 - c_1$ (where g_1 is high and c_1 is low), Player 1 will then choose to cheat. The game ends in the Nash equilibrium $\{Cheat, Help\}$ with payoffs $\{u_1 + g_1 - c_1, u_2 + g_2 - c_2\}$. However, if the costs of providing help are sufficiently high, as in the open-response tests, such that $g_2 < c_2$, Player 2 will refuse to help. In this case, Player 1's best response is not to cheat, resulting in the equilibrium $\{Not\ Cheat, Not\ Help\}$ with payoffs $\{u_1, u_2\}$. This is also the only possible equilibrium in the post-reform period, as competition offsets altruism ($g_2 = 0$).

Although simple, the model provides key insights into how cheating can occur between elite and non-elite students and offers testable predictions on how this behavior responds to variations in costs and shifts in incentives.

B Additional Figures

FIGURE B1 Examples of Cheating

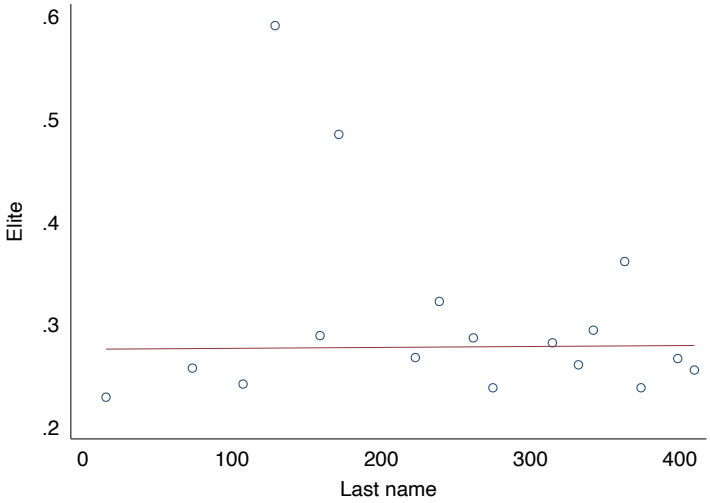
(A) Using Notes and Study Aids

(B) Obtaining Outside Help

(C) Copying Other Students

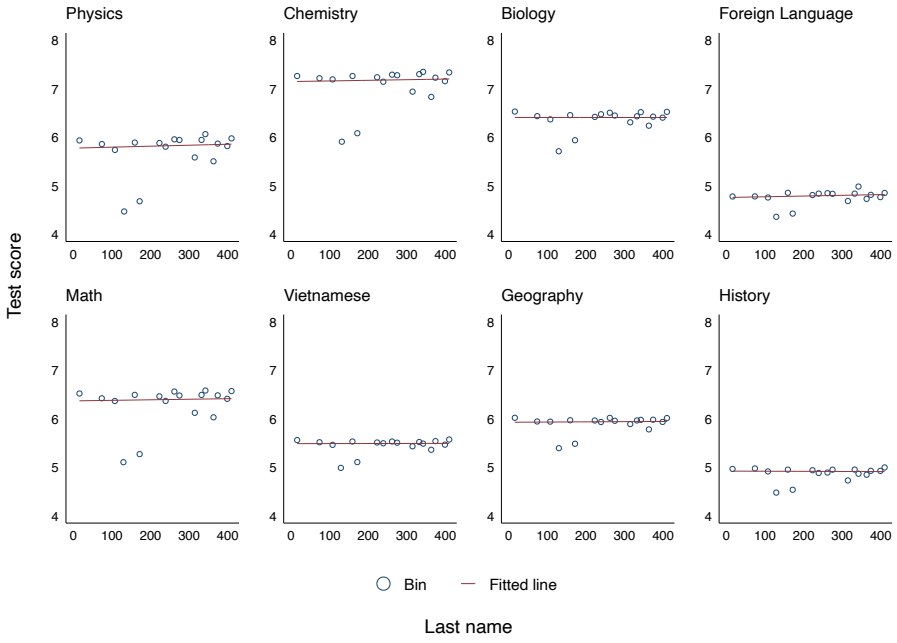
Notes: This figure illustrates several forms of cheating that have been reported to occur in the NHSE. Sources: Kiều 2012, Thanh and Quốc 2006, Quân and Hạ 2012.

FIGURE B2 Binned Scatterplot of Elite Status by Last Name



Notes: This figure shows the binned scatterplot of elite status among students by last name. The underlying sample includes all students who took the NHSE between 2002 and 2019.

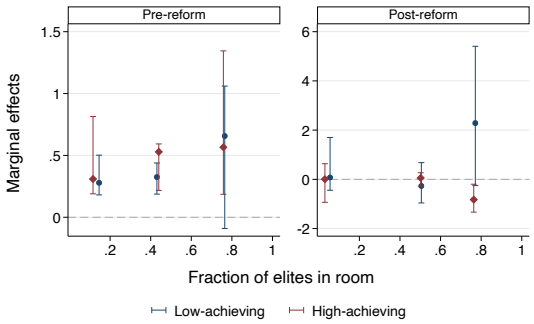
FIGURE B3 Binned Scatterplots of Test Scores by Last Name



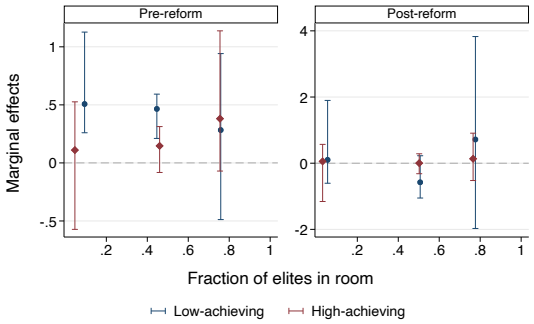
Notes: This figure shows the binned scatterplots of test scores among students by last name. The underlying sample includes all students who took the NHSE between 2002 and 2019.

FIGURE B4 Quantile Effects of Elite Density on Test Scores by Ability

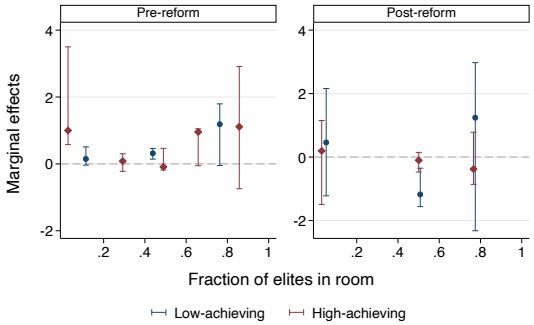
(A) Physics



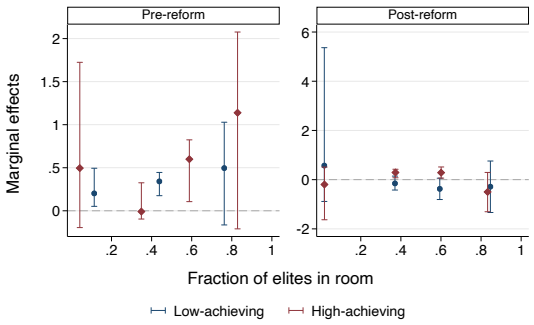
(B) Chemistry



(C) Biology



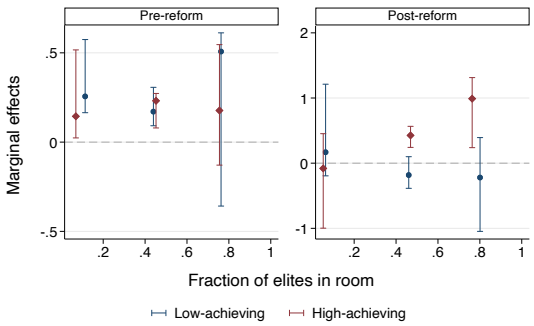
(D) Foreign Language



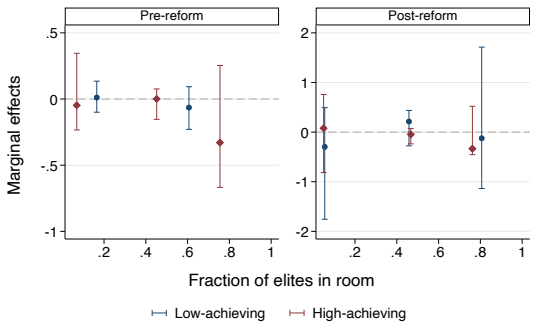
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2. The number of bins is chosen separately by period and baseline ability and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, and socioeconomic status). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE B4 Quantile Effects of Elite Density on Test Scores by Ability (Continued)

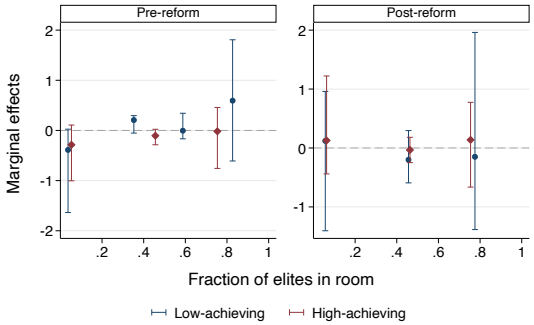
(E) Math



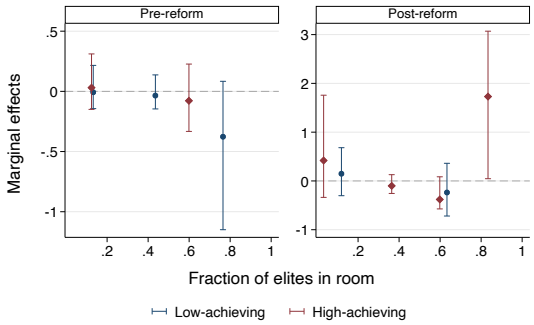
(F) Vietnamese



(G) Geography



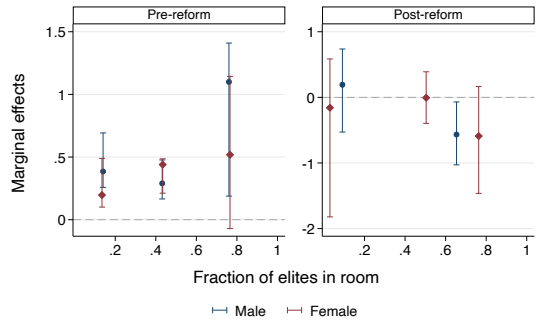
(H) History



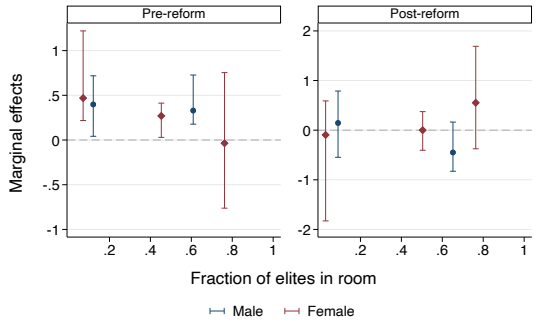
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2. The number of bins is chosen separately by period and baseline ability and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, and socioeconomic status). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE B5 Quantile Effects of Elite Density on Test Scores by Gender

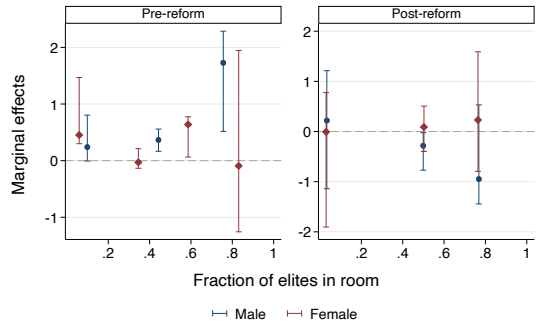
(A) Physics



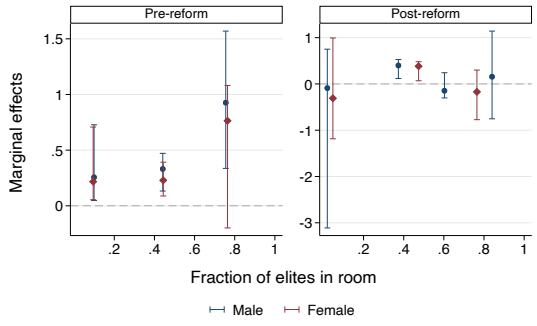
(B) Chemistry



(C) Biology



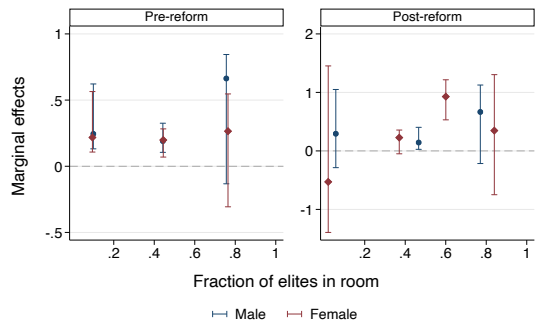
(D) Foreign Language



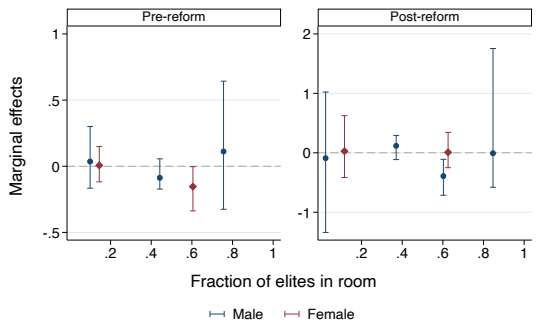
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2. The number of bins is chosen separately by period and gender and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE B5 Quantile Effects of Elite Density on Test Scores by Gender (Continued)

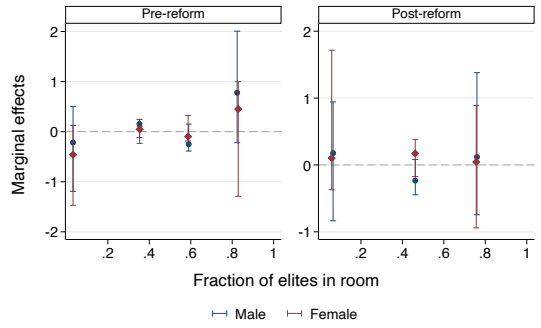
(E) Math



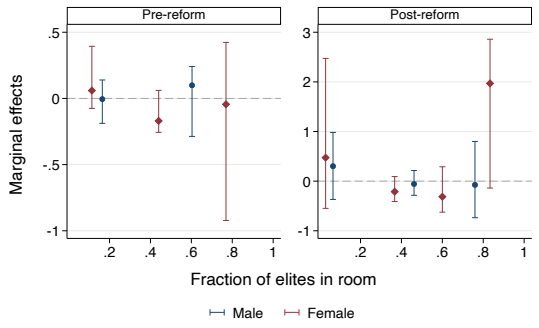
(F) Vietnamese



(G) Geography



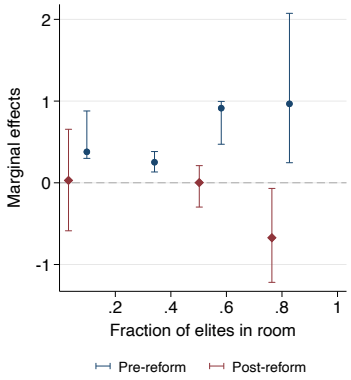
(H) History



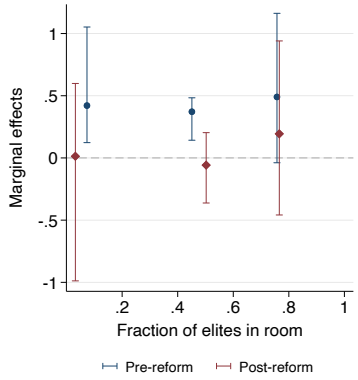
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2. The number of bins is chosen separately by period and gender and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE B6 **Robustness - No Covariates:** Quantile Effects of Elite Density on Test Scores

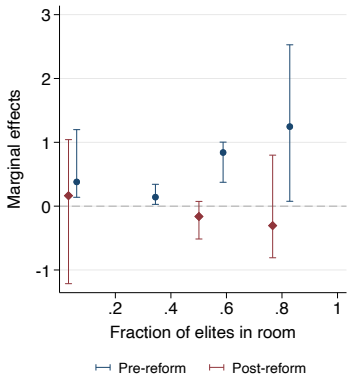
(A) Physics



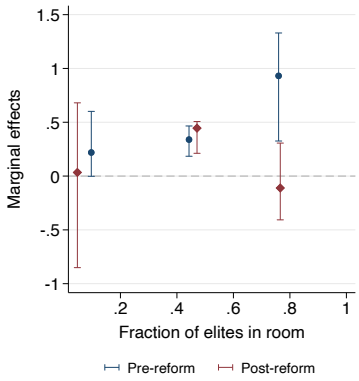
(B) Chemistry



(C) Biology



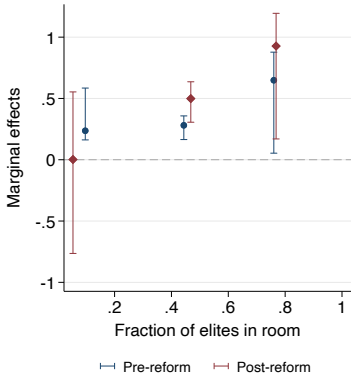
(D) Foreign Language



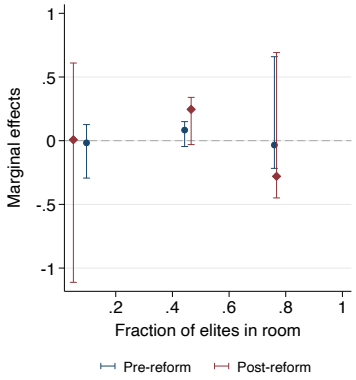
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2 but without covariates $\boldsymbol{X}_{ij}$. The number of bins is chosen separately by period and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE B6 **Robustness - No Covariates:** Quantile Effects of Elite Density on Test Scores (Continued)

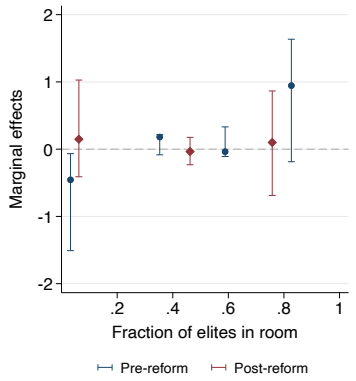
(E) Math



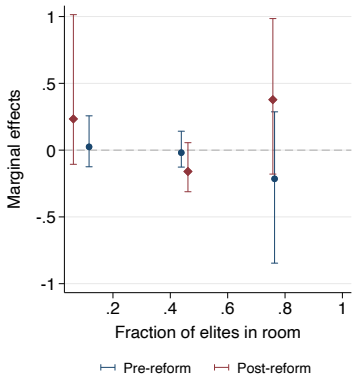
(F) Vietnamese



(G) Geography



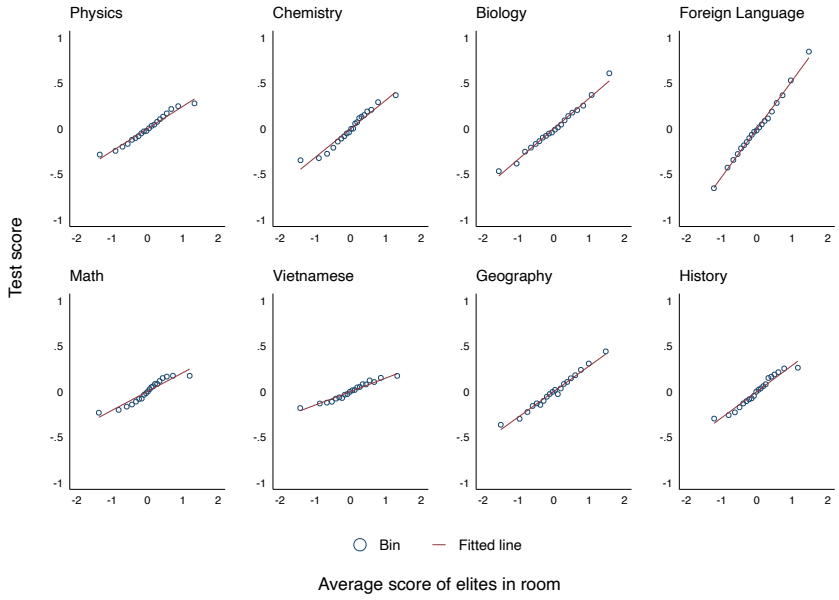
(H) History



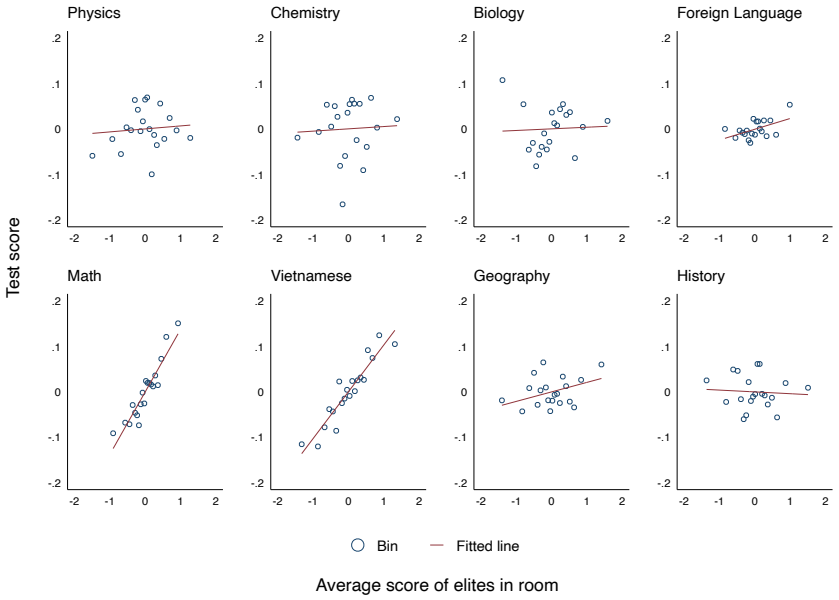
Notes: The pre-reform and post-reform periods span 2007-2013 and 2018-2019, respectively. The binned scatterplots present the effects of elite peer density on test scores over its distribution, as estimated from Equation 2 but without covariates $\boldsymbol{X}_{ij}$. The number of bins is chosen separately by period and is set to be IMSE-optimal. In each plot, a circle or diamond indicates a point estimate, while a spike illustrates the corresponding 95% confidence interval. The sample consists of non-elite students who took the NHSE in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level.

FIGURE B7 Binned Scatterplots of Test Scores of Non-elite Students against Elite Peers

(A) Pre-reform



(B) Post-reform



Notes: This figure shows the binned scatterplots of the test scores of non-elite students against the average test scores of their same-room elite peers. Both test scores are residualized on the share of elite students in a room, non-elite students' characteristics as included in the vector of covariates $\boldsymbol{X}$, and the fixed effects in Equation 1. The underlying samples include all non-elite students in the province who took the NHSE from 2007-2013 (Panel A) and from 2018-2019 (Panel B).

C Additional Tables

Table C1 CC Sample: Summary Statistics on Student Characteristics

A Pre-reform

	Non-elite			Elite		
	N	Mean	SD	*N*	Mean	SD
Background characteristics						
Female	20,322	0.50	0.50	3,923	0.51	0.50
Minority	20,322	0.05	0.21	3,923	0.00	0.07
No priority	20,322	0.85	0.35	3,923	0.89	0.31
Low priority	20,322	0.10	0.30	3,923	0.10	0.30
High priority	20,322	0.05	0.22	3,923	0.01	0.09
HS Academic standing: Low-achieving	20,322	0.46	0.50	3,923	0.01	0.12
HS Academic standing: High-achieving	20,322	0.54	0.50	3,923	0.99	0.12
Test performance						
Passing	20,322	0.92	0.27	3,923	0.99	0.08
Net final score	20,312	6.85	1.27	3,921	7.96	0.92
Gross final score	20,248	6.57	1.19	3,919	7.76	0.84
Physics	16,305	7.52	1.99	3,169	8.83	1.45
Chemistry	8,154	8.21	1.87	1,412	9.28	1.18
Biology	12,168	7.42	1.59	2,511	8.45	1.22
Foreign Language	20,322	6.40	1.99	3,923	7.57	1.83
Math	20,322	7.37	2.31	3,923	8.90	1.27
Vietnamese	20,322	5.31	1.57	3,923	6.52	1.42
Geography	12,105	5.72	1.24	2,522	6.62	1.17
History	12,234	5.08	2.04	2,155	6.11	1.76

B Post-reform

	Non-elite			Elite		
	N	Mean	SD	*N*	Mean	SD
Background characteristics						
Female	5,437	0.51	0.50	1,622	0.54	0.50
Minority	5,437	0.12	0.33	1,622	0.01	0.10
No priority	5,437	0.84	0.36	1,622	0.94	0.23
Low priority	5,437	0.03	0.18	1,622	0.04	0.19
High priority	5,437	0.12	0.33	1,622	0.02	0.13
HS Academic standing: Low-achieving	5,437	0.04	0.20	1,622	0.00	0.00
HS Academic standing: High-achieving	5,437	0.96	0.20	1,622	1.00	0.00
Test performance						
Passing	5,437	0.98	0.14	1,622	1.00	0.00
Net final score	5,393	5.32	1.40	1,622	6.39	0.83
Gross final score	5,122	5.24	0.99	1,622	6.37	0.82
Physics	2,553	5.07	1.77	1,041	6.29	1.80
Chemistry	2,553	4.64	1.69	1,041	6.39	1.82
Biology	2,553	4.07	1.23	1,041	4.86	1.71
Foreign Language	5,431	4.04	2.14	1,594	6.34	2.10
Math	5,437	5.53	1.73	1,622	7.26	1.32
Vietnamese	5,437	5.63	1.66	1,622	6.58	1.60
Geography	3,514	5.74	1.28	611	6.57	1.27
History	3,514	3.94	1.41	611	4.87	1.78

Notes: Panel A and Panel B summarize the pre-reform period (2007-2013) and the post-reform period (2018-2019), respectively. For each period, the sample consists of students who took the NHSE in the capital city in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). All test scores are on a 0-10 grading scale.

Table C2 CC Sample: Summary Statistics on Test Room Characteristics

A Pre-reform

		Physics	Chemistry	Biology	Foreign Language	Math	Vietnamese	Geography	History
No. HS with students in room	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	6.60	6.66	5.86	6.18	6.17	6.17	4.43	7.37
	SD	2.89	2.85	2.74	2.81	2.79	2.80	1.55	2.81
	Median	6.00	6.00	5.00	6.00	6.00	6.00	5.00	8.00
No. students in room	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	23.71	23.66	23.71	23.69	23.63	23.66	23.62	23.63
	SD	1.65	1.49	1.86	1.72	1.73	1.72	1.98	1.52
	Median	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00
$\mathbf{1}\{$Room has elite(s)$\}$	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	0.83	0.93	0.78	0.84	0.84	0.84	0.75	0.94
	SD	0.38	0.26	0.42	0.37	0.37	0.37	0.44	0.24
	Median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. elites in room	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	3.84	3.48	4.03	3.81	3.81	3.81	4.04	3.53
	SD	3.39	2.60	3.70	3.32	3.32	3.32	3.89	2.47
	Median	3.00	3.00	4.00	3.00	3.00	3.00	4.00	3.00
Frac. elites in room	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	0.17	0.16	0.18	0.17	0.17	0.17	0.18	0.16
	SD	0.15	0.12	0.17	0.15	0.15	0.15	0.18	0.12
	Median	0.13	0.13	0.17	0.14	0.14	0.14	0.17	0.13
Avg. score elites in room	*N*	686	377	486	863	863	863	466	574
	Mean	8.64	9.22	8.42	7.38	8.77	6.56	6.68	6.05
	SD	1.22	0.87	0.92	1.54	0.96	1.06	0.79	1.40
	Median	9.00	9.50	8.50	7.58	9.00	6.60	6.64	6.17
$\mathbf{1}\{$Room has subject elite(s)$\}$	*N*	686	377	486	863	863	863	466	574
	Mean	0.45	0.15	0.15	0.19	0.68	0.16	0.14	0.10
	SD	0.50	0.35	0.35	0.40	0.47	0.37	0.35	0.30
	Median	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
No. subject elites in room	*N*	686	377	486	863	863	863	466	574
	Mean	0.64	0.16	0.20	0.49	1.28	0.19	0.18	0.11
	SD	0.89	0.40	0.57	2.22	1.31	0.46	0.50	0.33
	Median	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Frac. subject elites in room	*N*	686	377	486	863	863	863	466	574
	Mean	0.03	0.01	0.01	0.02	0.06	0.01	0.01	0.00
	SD	0.04	0.02	0.03	0.12	0.06	0.02	0.02	0.01
	Median	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
$\mathbf{1}\{$Room has field elite(s)$\}$	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	0.70	0.72	0.65	0.31	0.68	0.31	0.27	0.35
	SD	0.46	0.45	0.48	0.46	0.47	0.46	0.44	0.48
	Median	1.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00
No. field elites in room	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	1.81	1.57	1.69	0.70	1.65	0.70	0.73	0.70
	SD	1.96	1.51	2.03	2.14	1.84	2.14	2.45	1.88
	Median	1.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00
Frac. field elites in room	*N*	826	406	624	1,030	1,030	1,030	624	610
	Mean	0.08	0.07	0.07	0.03	0.07	0.03	0.03	0.03
	SD	0.09	0.07	0.09	0.11	0.08	0.11	0.12	0.10
	Median	0.05	0.04	0.04	0.00	0.04	0.00	0.00	0.00

Notes: This table shows the student composition at the test room level. Panel A and Panel B summarize the pre-reform (2007-2013) and post-reform (2018-2019) period, respectively. For each period, the underlying sample consists of students who took the NHSE in the capital city in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). All test scores are on a 0-10 grading scale.

Table C2 CC Sample: Summary Statistics on Test Room Characteristics (Continued)

B Post-reform

		Physics	Chemistry	Biology	Foreign Language	Math	Vietnamese	Geography	History
No. HS with students in room	*N*	162	162	162	288	320	311	183	183
	Mean	2.94	3.06	3.01	3.44	3.90	3.67	3.85	3.96
	SD	0.60	0.80	0.73	0.82	1.35	1.01	1.01	1.13
	Median	3.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00
No. students in room	*N*	162	162	162	288	320	311	183	183
	Mean	22.63	22.72	22.51	23.50	23.39	23.41	22.80	23.15
	SD	2.78	2.67	2.89	1.39	1.79	1.68	2.51	2.23
	Median	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00
$\mathbf{1}\{\text{Room has elite(s)}\}$	*N*	162	162	162	288	320	311	183	183
	Mean	0.72	0.73	0.73	0.72	0.71	0.70	0.69	0.69
	SD	0.45	0.44	0.44	0.45	0.45	0.46	0.46	0.46
	Median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. elites in room	*N*	162	162	162	288	320	311	183	183
	Mean	6.45	6.48	6.48	5.55	5.19	5.24	3.36	3.36
	SD	5.41	5.39	5.39	4.65	4.38	4.43	3.06	3.06
	Median	6.00	6.00	6.00	6.00	5.00	6.00	3.00	3.00
Frac. elites in room	*N*	162	162	162	288	320	311	183	183
	Mean	0.30	0.30	0.30	0.25	0.23	0.23	0.15	0.15
	SD	0.24	0.24	0.24	0.21	0.19	0.19	0.14	0.14
	Median	0.30	0.30	0.30	0.26	0.24	0.26	0.13	0.13
Avg. score elites in room	*N*	116	119	119	207	228	218	127	127
	Mean	6.16	6.22	4.93	6.64	7.19	6.73	6.47	4.72
	SD	0.79	0.90	0.86	1.24	0.82	1.03	0.73	0.98
	Median	6.33	6.37	4.78	6.75	7.19	6.75	6.50	4.65
$\mathbf{1}\{\text{Room has subject elite(s)}\}$	*N*	116	117	117	207	226	218	126	126
	Mean	0.74	0.80	0.68	0.74	0.66	0.21	0.27	0.33
	SD	0.44	0.40	0.47	0.44	0.47	0.41	0.45	0.47
	Median	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00
No. subject elites in room	*N*	116	117	117	207	226	218	126	126
	Mean	1.84	2.21	1.12	2.06	1.64	0.32	0.48	0.54
	SD	1.53	1.85	1.14	2.56	1.78	0.71	0.92	0.94
	Median	2.00	2.00	1.00	1.00	1.00	0.00	0.00	0.00
Frac. subject elites in room	*N*	116	117	117	207	226	218	126	126
	Mean	0.08	0.10	0.05	0.09	0.07	0.01	0.02	0.02
	SD	0.07	0.08	0.05	0.13	0.08	0.03	0.04	0.04
	Median	0.09	0.09	0.04	0.05	0.04	0.00	0.00	0.00
$\mathbf{1}\{\text{Room has field elite(s)}\}$	*N*	162	162	162	288	320	311	183	183
	Mean	0.70	0.70	0.70	0.60	0.53	0.59	0.67	0.67
	SD	0.46	0.46	0.46	0.49	0.50	0.49	0.47	0.47
	Median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. field elites in room	*N*	162	162	162	288	320	311	183	183
	Mean	5.95	5.95	5.95	2.03	2.37	1.86	3.01	3.01
	SD	5.11	5.11	5.11	2.69	3.57	2.09	2.90	2.90
	Median	6.00	6.00	6.00	1.00	1.00	1.00	3.00	3.00
Frac. field elites in room	*N*	162	162	162	288	320	311	183	183
	Mean	0.27	0.27	0.27	0.09	0.10	0.08	0.14	0.14
	SD	0.23	0.23	0.23	0.13	0.16	0.09	0.13	0.13
	Median	0.26	0.26	0.26	0.04	0.04	0.04	0.13	0.13

Notes: This table shows the student composition at the test room level. Panel A and Panel B summarize the pre-reform (2007-2013) and post-reform (2018-2019) period, respectively. For each period, the underlying sample consists of students who took the NHSE in the capital city in the indicated years, were assigned to a room with at least another student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). All test scores are on a 0-10 grading scale.

Table C3 Average Effects of Elite Presence on Test Scores by Ability

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Low-achieving								
$\mathbf{1}\{\text{Room has elite(s)}\}$	0.099***	0.216***	0.053	−0.035	0.109***	−0.014	−0.139**	0.071**
	(0.032)	(0.077)	(0.040)	(0.033)	(0.034)	(0.027)	(0.058)	(0.032)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.336	0.353	0.193	0.363	0.460	0.335	0.195	0.503
Observations	93,586	69,008	77,419	134,975	134,370	134,608	78,062	86,308
B. High-achieving								
$\mathbf{1}\{\text{Room has elite(s)}\}$	0.236***	0.170**	0.209***	−0.049	0.130***	0.009	−0.142**	0.071
	(0.053)	(0.080)	(0.079)	(0.085)	(0.034)	(0.034)	(0.060)	(0.046)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.312	0.296	0.287	0.326	0.268	0.262	0.156	0.363
Observations	35,525	48,071	46,148	78,290	78,271	78,289	66,452	38,662

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Low-achieving								
$\mathbf{1}\{\text{Room has elite(s)}\}$	−0.024	0.023	−0.015	0.373***	0.162***	−0.050	−0.033	0.050
	(0.073)	(0.087)	(0.111)	(0.077)	(0.039)	(0.079)	(0.087)	(0.053)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.175	0.266	0.155	0.033	0.297	0.288	0.182	0.111
Observations	538	538	538	4,135	5,064	5,066	4,945	4,945
B. High-achieving								
$\mathbf{1}\{\text{Room has elite(s)}\}$	−0.065	−0.003	0.020	−0.067	0.034	−0.006	0.166	0.017
	(0.050)	(0.036)	(0.080)	(0.061)	(0.078)	(0.058)	(0.102)	(0.068)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.192	0.180	0.078	0.234	0.350	0.328	0.160	0.142
Observations	12,122	12,122	12,122	35,690	37,291	37,289	26,513	26,513

Notes: This table presents the results of estimating Equation 1 where the main explanatory variable is the presence of elite students in a non-elite student's test room, separately by ability. Panel A reports the estimates for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). The sample consists of non-elite students who took the NHSE in the indicated years and had complete information on observable demographic characteristics (gender, minority ethnicity, and socioeconomic status). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C4 Average Effects of Elite Density on Test Scores by Ability

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Low-achieving								
Frac. of elites in room	0.301***	0.484***	0.236***	0.262***	0.229***	0.002	0.003	−0.023
	(0.043)	(0.071)	(0.056)	(0.045)	(0.039)	(0.035)	(0.061)	(0.046)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.336	0.354	0.194	0.363	0.461	0.335	0.194	0.503
Observations	93,586	69,008	77,419	134,975	134,370	134,608	78,062	86,308
B. High-achieving								
Frac. of elites in room	0.392***	0.138**	0.333***	0.275***	0.181***	−0.036	−0.192***	0.017
	(0.064)	(0.070)	(0.080)	(0.076)	(0.037)	(0.041)	(0.058)	(0.062)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.312	0.296	0.287	0.326	0.268	0.262	0.156	0.363
Observations	35,525	48,071	46,148	78,290	78,271	78,289	66,452	38,662

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Low-achieving								
Frac. of elites in room	−0.038	−0.194	−0.238	−0.155	−0.042	−0.024	−0.049	0.067
	(0.129)	(0.131)	(0.154)	(0.101)	(0.089)	(0.140)	(0.158)	(0.110)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.175	0.267	0.157	0.034	0.297	0.288	0.182	0.111
Observations	538	538	538	4,135	5,064	5,066	4,945	4,945
B. High-achieving								
Frac. of elites in room	−0.080	0.034	−0.063	0.199***	0.305***	−0.026	0.049	0.017
	(0.107)	(0.115)	(0.131)	(0.063)	(0.066)	(0.062)	(0.080)	(0.074)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.192	0.180	0.078	0.235	0.350	0.328	0.160	0.142
Observations	12,122	12,122	12,122	35,690	37,291	37,289	26,513	26,513

Notes: This table presents the results of estimating Equation 1 where the main explanatory variable is the density of elite students in a non-elite student's test room, separately by ability. Panel A reports the estimates for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). The sample consists of non-elite students who took the NHSE in the indicated years and had complete information on observable demographic characteristics (gender, minority ethnicity, and socioeconomic status). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C5 Average Effects of Elite Presence on Test Scores by Gender

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
$\mathbf{1}\{\text{Room has elite(s)}\}$	0.124***	0.137	0.065	−0.064	0.102***	−0.027	−0.096*	0.038
	(0.046)	(0.100)	(0.061)	(0.051)	(0.039)	(0.033)	(0.057)	(0.037)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.419	0.403	0.264	0.406	0.523	0.345	0.256	0.522
Observations	62,117	56,468	58,318	101,711	101,320	101,467	67,014	60,865
B. Female								
$\mathbf{1}\{\text{Room has elite(s)}\}$	0.142***	0.304***	0.132***	−0.012	0.128***	0.024	−0.144***	0.100***
	(0.033)	(0.065)	(0.046)	(0.046)	(0.031)	(0.026)	(0.051)	(0.036)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.457	0.441	0.295	0.427	0.579	0.372	0.286	0.520
Observations	66,994	60,611	65,249	111,554	111,321	111,430	77,500	64,105

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
$\mathbf{1}\{\text{Room has elite(s)}\}$	−0.009	−0.055	−0.020	−0.074	0.078	−0.058	0.075	−0.020
	(0.052)	(0.051)	(0.081)	(0.159)	(0.082)	(0.065)	(0.089)	(0.056)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.276	0.251	0.110	0.231	0.462	0.310	0.229	0.162
Observations	6,786	6,786	6,786	18,802	20,545	20,546	14,866	14,866
B. Female								
$\mathbf{1}\{\text{Room has elite(s)}\}$	−0.120*	0.055	0.007	−0.105	0.154**	0.085	0.138	0.149**
	(0.065)	(0.070)	(0.068)	(0.095)	(0.069)	(0.092)	(0.109)	(0.072)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.183	0.208	0.077	0.221	0.345	0.231	0.188	0.173
Observations	5,874	5,874	5,874	21,023	21,810	21,809	16,592	16,592

Notes: This table presents the results of estimating Equation 1 where the main explanatory variable is the presence of elite students in a non-elite student's test room, separately by gender. Panel A reports the estimates for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). The sample consists of non-elite students who took the NHSE in the indicated years and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C6 Average Effects of Elite Density on Test Scores by Gender

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Frac. of elites in room	0.365***	0.388***	0.328***	0.302***	0.235***	−0.009	−0.018	0.005
	(0.054)	(0.078)	(0.068)	(0.059)	(0.042)	(0.041)	(0.062)	(0.054)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.419	0.404	0.265	0.407	0.523	0.345	0.256	0.522
Observations	62,117	56,468	58,318	101,711	101,320	101,467	67,014	60,865
B. Female								
Frac. of elites in room	0.290***	0.366***	0.235***	0.236***	0.211***	−0.015	−0.121**	−0.026
	(0.047)	(0.072)	(0.063)	(0.053)	(0.039)	(0.036)	(0.058)	(0.052)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.457	0.441	0.295	0.427	0.579	0.372	0.285	0.520
Observations	66,994	60,611	65,249	111,554	111,321	111,430	77,500	64,105

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Frac. of elites in room	−0.096	−0.081	−0.202	0.170**	0.234***	−0.051	−0.022	0.113
	(0.137)	(0.139)	(0.141)	(0.074)	(0.073)	(0.073)	(0.099)	(0.089)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.276	0.251	0.110	0.231	0.462	0.310	0.229	0.162
Observations	6,786	6,786	6,786	18,802	20,545	20,546	14,866	14,866
B. Female								
Frac. of elites in room	−0.123	0.044	0.080	0.132	0.299***	0.022	0.134	−0.019
	(0.143)	(0.143)	(0.179)	(0.086)	(0.081)	(0.084)	(0.101)	(0.095)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.183	0.208	0.077	0.221	0.345	0.231	0.188	0.173
Observations	5,874	5,874	5,874	21,023	21,810	21,809	16,592	16,592

Notes: This table presents the results of estimating Equation 1 where the main explanatory variable is the density of elite students in a non-elite student's test room, separately by gender. Panel A reports the estimates for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). The sample consists of non-elite students who took the NHSE in the indicated years and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C7 CC Sample: Relationship between Elite Quality and Non-elite Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	0.205***	0.185***	0.319***	0.434***	0.084***	0.090***	0.264***	0.203***
	(0.020)	(0.022)	(0.028)	(0.020)	(0.014)	(0.012)	(0.025)	(0.022)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.372	0.266	0.320	0.444	0.496	0.440	0.276	0.486
Observations	13,046	7,493	8,959	16,454	16,397	16,426	8,431	11,376

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	0.025	0.028	0.036	0.015	0.006	0.112***	0.031	0.002
	(0.021)	(0.029)	(0.029)	(0.021)	(0.018)	(0.023)	(0.021)	(0.024)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.330	0.175	0.080	0.349	0.579	0.375	0.160	0.096
Observations	1,572	1,599	1,599	3,230	3,538	3,444	2,328	2,328

Notes: This table presents the relationship between the quality of same-room elite peers and non-elite students' test scores. Panel A reports the regression results for the pre-reform period (2007-2013) while Panel B reports those for the post-reform period (2018-2019). All columns control for the density of same-room elite peers. The sample consists of non-elite students who took the NHSE in the capital city in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C8 Reduced-form Effects of Subject Elite Density on Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of subject elites in room	1.804***	−3.297***	6.020***	5.931***	−0.117	−0.457	0.273	0.022
	(0.431)	(1.249)	(1.091)	(0.891)	(0.235)	(0.461)	(0.903)	(1.027)
Frac. of elites in room	0.194	0.039	0.454*	0.629***	0.083	−0.012	−0.242	0.064
	(0.181)	(0.253)	(0.237)	(0.214)	(0.135)	(0.103)	(0.192)	(0.200)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.346	0.238	0.244	0.333	0.490	0.435	0.232	0.463
Observations	13,046	7,470	8,959	16,431	16,374	16,403	8,408	11,353

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of subject elites in room	−0.736**	−0.080	−0.171	0.247	0.410*	1.570*	−0.125	0.650
	(0.335)	(0.552)	(0.576)	(0.230)	(0.231)	(0.808)	(0.775)	(1.123)
Frac. of elites in room	0.268	−0.066	−0.016	0.050	−0.144	0.069	0.111	0.433*
	(0.252)	(0.303)	(0.327)	(0.143)	(0.139)	(0.163)	(0.235)	(0.245)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.331	0.174	0.074	0.349	0.579	0.370	0.149	0.096
Observations	1,572	1,577	1,577	3,230	3,514	3,444	2,312	2,312

Notes: This table presents the reduced form results of estimating the IV model characterized by Equations 4-5, using the share of subject-specific elites in a test room as an instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C9 **First Stage:** Subject of Specialization and Overall Performance among Elites by Gender

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Frac. of subject elites in room	4.167***	1.237	12.815***	13.225***	1.353*	9.832***	−0.329	5.188**
	(0.681)	(2.783)	(2.081)	(1.770)	(0.717)	(1.530)	(2.118)	(2.317)
Frac. of elites in room	0.253	1.352*	1.216**	0.212	0.088	−0.501	0.209	0.641
	(0.400)	(0.756)	(0.489)	(0.357)	(0.447)	(0.360)	(0.462)	(0.470)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	37.403	0.198	37.915	55.798	3.561	41.266	0.024	5.014
Adj. R^2	0.295	0.166	0.206	0.394	0.227	0.396	0.350	0.465
Observations	6,600	3,716	4,531	8,248	8,212	8,227	4,232	5,629
B. Female								
Frac. of subject elites in room	3.862***	4.881*	14.493***	11.423***	1.362	7.569***	2.470	4.903*
	(1.289)	(2.803)	(2.197)	(1.339)	(0.840)	(1.492)	(1.947)	(2.571)
Frac. of elites in room	0.877*	1.209	−0.066	−0.071	0.768	−0.592	−0.798	0.056
	(0.514)	(0.787)	(0.567)	(0.379)	(0.513)	(0.382)	(0.485)	(0.524)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	8.976	3.034	43.528	72.810	2.627	25.753	1.609	3.637
Adj. R^2	0.297	0.125	0.198	0.359	0.214	0.327	0.337	0.443
Observations	6,446	3,754	4,428	8,183	8,162	8,176	4,176	5,724

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Frac. of subject elites in room	4.476***	5.847***	5.916***	5.318***	3.256***	10.058***	10.491***	10.835**
	(1.306)	(1.147)	(1.644)	(0.730)	(0.791)	(1.900)	(2.933)	(4.571)
Frac. of elites in room	−0.089	−0.344	−1.214	−1.307**	−1.187*	0.292	0.511	0.284
	(0.889)	(0.696)	(0.738)	(0.518)	(0.616)	(0.498)	(1.047)	(0.965)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	11.743	26.002	12.949	53.068	16.931	28.030	12.793	5.618
Adj. R^2	0.417	0.419	0.373	0.681	0.618	0.496	0.205	0.320
Observations	933	938	938	1,597	1,795	1,741	1,058	1,058
B. Female								
Frac. of subject elites in room	6.853***	3.972**	5.896***	3.184***	4.608***	9.863***	7.428***	6.019*
	(1.407)	(1.526)	(1.277)	(0.610)	(0.788)	(1.636)	(1.978)	(3.095)
Frac. of elites in room	0.223	−0.133	−1.940**	−2.382***	−0.220	−0.793	0.924	1.146
	(1.017)	(1.028)	(0.787)	(0.393)	(0.541)	(0.599)	(0.835)	(0.913)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	23.718	6.778	21.330	27.260	34.160	36.330	14.098	3.783
Adj. R^2	0.480	0.197	0.563	0.669	0.612	0.444	0.327	0.430
Observations	639	639	639	1,633	1,719	1,703	1,254	1,254

Notes: This table presents the first stage results of estimating the IV model characterized by Equations 4-5 separately by gender, using the share of subject-specific elites in a test room as an instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C10 **Second Stage:** Effects of Elite Quality on Test Scores by Gender

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Avg. score of elites in room	0.502***	−2.371	0.434***	0.598***	−0.351	−0.002	−1.030	0.116
	(0.116)	(6.128)	(0.091)	(0.064)	(0.314)	(0.074)	(8.535)	(0.220)
Frac. of elites in room	0.205	3.411	0.281	0.557***	0.245	−0.076	0.008	−0.033
	(0.275)	(8.494)	(0.346)	(0.188)	(0.270)	(0.137)	(1.754)	(0.270)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.003	−7.191	0.092	0.226	−0.119	0.063	−1.217	0.065
Observations	6,600	3,716	4,531	8,248	8,212	8,227	4,232	5,629
B. Female								
Avg. score of elites in room	0.312**	−0.715	0.449***	0.413***	0.113	−0.095	0.213	−0.104
	(0.151)	(0.638)	(0.083)	(0.066)	(0.198)	(0.077)	(0.441)	(0.309)
Frac. of elites in room	−0.298	0.871	0.087	0.558**	−0.172	0.046	−0.026	0.188
	(0.290)	(1.108)	(0.304)	(0.219)	(0.272)	(0.145)	(0.424)	(0.307)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.050	−0.810	0.181	0.161	0.144	0.057	0.102	−0.005
Observations	6,446	3,754	4,428	8,183	8,162	8,176	4,176	5,724

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Avg. score of elites in room	−0.220**	0.070	0.061	0.113	−0.040	0.039	−0.013	0.184
	(0.106)	(0.112)	(0.098)	(0.072)	(0.086)	(0.103)	(0.126)	(0.210)
Frac. of elites in room	0.402	−0.235	−0.204	0.171	0.051	0.138	−0.134	0.074
	(0.338)	(0.336)	(0.311)	(0.201)	(0.149)	(0.178)	(0.336)	(0.433)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.060	−0.014	−0.014	−0.013	0.019	0.011	0.002	−0.039
Observations	933	938	938	1,597	1,795	1,741	1,058	1,058
B. Female								
Avg. score of elites in room	−0.062	−0.256	−0.149	0.048	0.206***	0.230**	−0.004	−0.045
	(0.096)	(0.226)	(0.133)	(0.092)	(0.078)	(0.090)	(0.138)	(0.209)
Frac. of elites in room	−0.066	0.230	0.309	0.138	−0.302	0.174	0.289	0.685
	(0.498)	(0.538)	(0.578)	(0.220)	(0.211)	(0.256)	(0.399)	(0.483)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.037	−0.077	−0.037	−0.012	−0.047	0.002	−0.010	−0.005
Observations	639	639	639	1,633	1,719	1,703	1,254	1,254

Notes: This table presents the second stage results of estimating the IV model characterized by Equations 4-5 separately by gender, using the share of subject-specific elites in a test room as an instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C11 **Alternative IV - First Stage:** Field of Specialization and Overall Performance among Elites

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of field elites in room	5.880***	3.668***	4.688***	7.726***	1.872***	6.528***	1.073	0.736
	(0.558)	(1.100)	(0.648)	(0.683)	(0.494)	(0.710)	(0.953)	(0.822)
Frac. of elites in room	−1.951***	−0.381	−0.753	−0.276	−0.021	−1.167***	−0.330	0.242
	(0.430)	(0.774)	(0.573)	(0.321)	(0.406)	(0.316)	(0.400)	(0.437)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	111.018	11.117	52.309	127.886	14.343	84.484	1.268	0.800
Adj. R^2	0.370	0.155	0.187	0.389	0.212	0.404	0.361	0.441
Observations	13,046	7,493	8,959	16,454	16,397	16,426	8,431	11,376

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Frac. of field elites in room	0.872	1.090	1.345	2.952***	2.452***	5.960***	4.621*	5.710**
	(1.952)	(2.255)	(2.027)	(0.534)	(0.390)	(0.534)	(2.425)	(2.331)
Frac. of elites in room	−0.240	−0.232	−1.701	−2.115***	−0.848	−1.845***	−2.364	−3.546*
	(2.005)	(2.332)	(2.105)	(0.438)	(0.531)	(0.443)	(2.327)	(1.980)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	0.200	0.234	0.440	30.623	39.483	124.587	3.633	6.000
Adj. R^2	0.367	0.299	0.364	0.649	0.606	0.607	0.273	0.375
Observations	1,572	1,599	1,599	3,230	3,538	3,444	2,328	2,328

Notes: This table presents the first stage results of estimating the IV model characterized by Equations 4-5, using the share of field-specific elites in a test room as an alternative instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C12 **Alternative IV - Second Stage:** Effects of Elite Quality on Test Scores

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	0.263***	0.339**	0.330***	0.544***	0.013	−0.049	−0.066	0.089
	(0.041)	(0.132)	(0.068)	(0.055)	(0.103)	(0.039)	(0.522)	(0.527)
Frac. of elites in room	0.160	−0.511	0.359	0.554***	0.037	−0.043	−0.246	0.032
	(0.174)	(0.319)	(0.233)	(0.167)	(0.145)	(0.105)	(0.238)	(0.270)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.099	0.055	0.155	0.226	0.126	0.167	0.025	0.072
Observations	13,046	7,493	8,959	16,454	16,397	16,426	8,431	11,376

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
Avg. score of elites in room	−0.439	0.136	0.615	0.338***	0.663***	0.275***	0.041	0.124
	(1.074)	(0.633)	(1.032)	(0.097)	(0.137)	(0.045)	(0.127)	(0.129)
Frac. of elites in room	0.436	−0.187	0.247	0.417**	−0.244	0.085	0.023	0.310
	(0.806)	(0.553)	(0.728)	(0.201)	(0.353)	(0.194)	(0.333)	(0.316)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.195	−0.020	−0.226	−0.053	−0.433	0.158	0.004	−0.010
Observations	1,572	1,599	1,599	3,230	3,538	3,444	2,328	2,328

Notes: This table presents the second stage results of estimating the IV model characterized by Equations 4-5, using the share of field-specific elites in a test room as an alternative instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (gender, minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C13 **Alternative IV - First Stage:** Field of Specialization and Overall Performance among Elites by Gender

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Frac. of field elites in room	5.232***	2.296*	3.814***	9.299***	1.660***	7.240***	1.431	1.507
	(0.577)	(1.374)	(0.752)	(0.854)	(0.595)	(0.946)	(1.167)	(1.017)
Frac. of elites in room	−1.982***	0.188	−0.010	−0.164	−0.322	−0.988***	0.087	0.573
	(0.455)	(1.041)	(0.646)	(0.366)	(0.514)	(0.360)	(0.461)	(0.483)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	82.140	2.795	25.711	118.616	7.780	58.605	1.503	2.196
Adj. R^2	0.346	0.175	0.196	0.404	0.231	0.427	0.371	0.463
Observations	6,600	3,728	4,531	8,260	8,224	8,239	4,244	5,641
B. Female								
Frac. of field elites in room	6.505***	5.192***	5.390***	7.020***	1.984***	6.084***	1.064	0.426
	(0.745)	(1.461)	(0.793)	(0.797)	(0.629)	(0.790)	(1.133)	(0.993)
Frac. of elites in room	−1.786***	−0.732	−1.274*	−0.374	0.322	−1.261***	−0.879*	0.108
	(0.557)	(0.845)	(0.678)	(0.416)	(0.510)	(0.399)	(0.499)	(0.599)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	76.268	12.630	46.214	77.481	9.943	59.355	0.882	0.184
Adj. R^2	0.376	0.159	0.198	0.340	0.220	0.364	0.355	0.439
Observations	6,446	3,765	4,428	8,194	8,173	8,187	4,187	5,735

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Frac. of field elites in room	0.681	0.023	0.542	4.032***	2.590***	5.961***	5.102**	5.393**
	(2.005)	(2.057)	(2.193)	(0.743)	(0.453)	(0.645)	(2.471)	(2.466)
Frac. of elites in room	−0.210	0.849	−0.966	−1.445***	−1.579**	−1.068**	−2.691	−3.399
	(2.092)	(2.169)	(2.304)	(0.532)	(0.631)	(0.439)	(2.440)	(2.205)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	0.115	0.000	0.061	29.428	32.768	85.283	4.265	4.784
Adj. R^2	0.359	0.353	0.303	0.661	0.638	0.591	0.258	0.312
Observations	933	948	948	1,597	1,807	1,741	1,072	1,072
B. Female								
Frac. of field elites in room	1.357	2.453	2.241	2.539***	2.388***	5.998***	3.599	5.695*
	(2.250)	(2.788)	(2.283)	(0.460)	(0.465)	(0.580)	(2.973)	(3.002)
Frac. of elites in room	−0.415	−1.563	−2.639	−2.548***	−0.054	−2.753***	−1.494	−3.393
	(2.167)	(2.776)	(2.148)	(0.416)	(0.536)	(0.546)	(2.694)	(2.367)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	0.364	0.774	0.963	30.478	26.320	106.969	1.465	3.599
Adj. R^2	0.374	0.238	0.484	0.659	0.588	0.577	0.285	0.433
Observations	639	651	651	1,633	1,731	1,703	1,256	1,256

Notes: This table presents the first stage results of estimating the IV model characterized by Equations 4-5 separately by gender, using the share of field-specific elites in a test room as an alternative instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C14 **Alternative IV - Second Stage:** Effects of Elite Quality on Test Scores by Gender

A Pre-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Avg. score of elites in room	0.265***	0.306	0.264**	0.649***	−0.142	−0.001	0.302	0.016
	(0.058)	(0.259)	(0.104)	(0.060)	(0.170)	(0.052)	(0.368)	(0.354)
Frac. of elites in room	0.436**	−0.259	0.617**	0.518***	0.132	−0.075	−0.257	0.045
	(0.212)	(0.446)	(0.302)	(0.189)	(0.185)	(0.135)	(0.221)	(0.368)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.072	0.049	0.115	0.213	0.053	0.063	0.093	0.033
Observations	6,600	3,728	4,531	8,260	8,224	8,239	4,244	5,641
B. Female								
Avg. score of elites in room	0.257***	0.299**	0.374***	0.452***	0.092	−0.100*	−0.356	0.328
	(0.048)	(0.129)	(0.072)	(0.070)	(0.115)	(0.052)	(0.952)	(1.096)
Frac. of elites in room	−0.229	−0.567	0.160	0.522**	−0.148	0.045	−0.435	0.100
	(0.231)	(0.417)	(0.298)	(0.224)	(0.197)	(0.145)	(0.825)	(0.335)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.062	0.037	0.192	0.156	0.145	0.055	−0.211	0.056
Observations	6,446	3,765	4,428	8,194	8,173	8,187	4,187	5,735

B Post-reform

	(1) Physics	(2) Chemistry	(3) Biology	(4) Foreign Language	(5) Math	(6) Vietnamese	(7) Geography	(8) History
A. Male								
Avg. score of elites in room	−0.980	−39.755	1.011	0.288***	0.626***	0.157***	0.192	0.074
	(2.761)	(3590.724)	(4.088)	(0.101)	(0.157)	(0.045)	(0.162)	(0.185)
Frac. of elites in room	0.723	34.440	0.235	0.229	0.204	0.059	−0.460	0.199
	(1.516)	(3131.997)	(1.998)	(0.248)	(0.375)	(0.180)	(0.526)	(0.390)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.933	−1271.727	−0.808	−0.055	−0.385	0.009	−0.023	−0.016
Observations	933	948	948	1,597	1,807	1,741	1,072	1,072
B. Female								
Avg. score of elites in room	0.075	0.709	0.634	0.390***	0.647***	0.378***	−0.201	0.229
	(0.598)	(0.955)	(0.830)	(0.129)	(0.174)	(0.067)	(0.369)	(0.219)
Frac. of elites in room	−0.177	−0.364	0.822	0.567**	−0.744*	0.249	0.654	0.163
	(0.712)	(1.030)	(1.075)	(0.282)	(0.413)	(0.299)	(0.741)	(0.490)
Test center × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High school × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	−0.026	−0.506	−0.209	−0.084	−0.406	−0.034	−0.069	−0.032
Observations	639	651	651	1,633	1,731	1,703	1,256	1,256

Notes: This table presents the second stage results of estimating the IV model characterized by Equations 4-5 separately by gender, using the share of field-specific elites in a test room as an alternative instrument. Panel A and Panel B report the estimates for the pre-reform (2007-2011) and post-reform (2018-2019) period, respectively. The sample consists of non-elite students in the capital city who took the NHSE in the indicated years, were assigned to a room with at least one elite student, and had complete information on observable demographic characteristics (minority ethnicity, socioeconomic status, and academic standing in grade 12). Test scores are standardized within each period using the entire sample. Standard errors are clustered at the test room × year level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.