Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Persistent legacy of the 1075–1919 Vietnamese imperial examinations in contemporary quantity and quality of education

Tien Manh Vu、Hiroyuki Yamada

3 June, 2020
DP2020-012
https://ies.keio.ac.jp/en/publications/13106/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
3 June, 2020

Persistent legacy of the 1075–1919 Vietnamese imperial examinations in contemporary quantity and quality of education

Tien Manh Vu、 Hiroyuki Yamada

Keio-IES DP2020-012
3 June, 2020
JEL Classification: I25, N35, O15
Keywords: Education; Human Capital; Imperial Examination; Historical Legacy; Vietnam

Abstract

We investigated the impact of individuals who passed the Vietnamese imperial examinations (1075-1919) on the present-day quantity and quality of education in their home districts. We layered the 2009 Population and Housing Census and the 2009 National Entrance Exams to University (NEEU) test scores on the geographical distribution of imperial test takers' home districts. We constructed a novel instrumental variable representing the average distance between the examinees' home districts and the corresponding imperial examination venues. We found a persistent legacy in the average years of schooling, literacy rate, school attendance rate, NEEU test scores, and primary school dropout rate.

Tien Manh Vu
Asian Growth Research Institute
11-4 Otemachi, Kokura-kita, Kitakyushu, Fukuoka
vu@agi.or.jp

Hiroyuki Yamada
Faculty of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo
hyamada@econ.keio.ac.jp

Acknowledgement : This work was supported by JSPS (Japan Society for the Promotion of Science) KAKENHI Grant Numbers 18K12784, 18K01580, 19H00619, and 20H01506, and project grants from the Asian Growth Research Institute, Kyushu University, and Keio Gijuku Academic Development Funds.

Persistent legacy of the 1075–1919 Vietnamese imperial examinations in contemporary quantity and quality of education

Tien Manh Vu[†] and Hiroyuki Yamada[††]

Abstract

We investigated the impact of individuals who passed the Vietnamese imperial examinations (1075–1919) on the present-day quantity and quality of education in their home districts. We layered the 2009 Population and Housing Census and the 2009 National Entrance Exams to University (NEEU) test scores on the geographical distribution of imperial test takers' home districts. We constructed a novel instrumental variable representing the average distance between the examinees' home districts and the corresponding imperial examination venues. We found a persistent legacy in the average years of schooling, literacy rate, school attendance rate, NEEU test scores, and primary school dropout rate.

Keywords: Education; Human Capital; Imperial Examination; Historical Legacy; Vietnam

JEL classification: I25, N35, O15

Acknowledgements

This work was supported by JSPS (Japan Society for the Promotion of Science) KAKENHI Grant Numbers 18K12784, 18K01580, 19H00619, and 20H01506, and project grants from the Asian Growth Research Institute, Kyushu University, and Keio Gijuku Academic Development Funds.

[†] Corresponding author. Asian Growth Research Institute, Osaka University, and Kyushu University. 11–4 Otemachi, Kokura-kita, Kitakyushu, Fukuoka 803–0814, Japan.
Tel.: +81 93 583 6202, Fax: +81 93 583 6576.

E-mail: vu@agi.or.jp.

[††] Faculty of Economics, Keio University.
2–15–45 Mita, Minato–ku, Tokyo 108–8345, Japan.

E-mail: hyamada@econ.keio.ac.jp.

1. Introduction

Some historical legacies have been advantageous for regional development while others have been disadvantageous (Nunn, 2020). The mechanisms include formal institutions (Acemoglu et al., 2001), geography, historical events, and informal institutions, that is, cultural norms, which Becker et al. (2016) specified as kinship, schooling, and social interactions. However, to our knowledge, literature on the economic rationale of informal institutional effects is scarce.

A culture of learning and determination to learn may be an aspect of social interaction. For example, for centuries, a popular Dong Ho folk woodcut print that pays tribute to those who were awarded the highest rank in the imperial examinations (*Vinh quy bái tổ* in Vietnamese, and hereafter, "paying tribute") has been purchased during the Vietnamese New Year holidays as a sign of luck. It is usually hung conspicuously indoors to motivate young family members to study hard and achieve academic success. Even 91 years after the last Vietnamese imperial examination, the theme still reverberates throughout Vietnamese society. In 2010, shortly after receiving the Fields Medal for proving the fundamental lemma for automorphic forms, Professor Ngô Bảo Châu, a Vietnamese-French mathematician, returned to his ancestral cemetery in Ba Dinh district, Hanoi to worship, report his success, and pay tribute to his ancestor (Thuan Hoa, 2010).

This study seeks empirical evidence of whether a culture of learning has passed from generation to generation in Vietnam and attempts to find the economic rationale behind the statistical results. Specifically, we investigate whether an association exists between the success of the test takers who passed the 1075–1919 imperial examinations at the national level[1] (hereafter, "imperial elites") and the educational achievements of people living in the district today that was home to these imperial elites. Our analysis examines educational outcomes at two levels, namely the district aggregated level (covering 687 districts[2]) and the individual level (covering about 0.68 million test takers born in 1991). In each district, we consider both the number of imperial elites and the density of imperial elites per square

[1] We refer to *Kỳ thi Đại khoa (thi Hội)* in Vietnamese as imperial examinations. Heavily influenced by the Chinese imperial examinations, the Vietnamese imperial examinations lasted 15 years longer. The examinations were used to select public officials to work in the imperial government, either at the central or the local level. The written examinations were conducted using mainly Chinese characters. See Appendix 1 for further details.

[2] There were 689 districts in the 2009 Population and Housing Census (100% version).

kilometer (as an alternative). We apply an instrumental variable (IV) approach for our analysis of both levels.

Our study is also motivated by voids in the literature and specific facts. First, to our knowledge, empirical evidence of a persistent culture of learning in Vietnam remains scant. Second, to pass the imperial examinations, candidates had to devote their efforts to mastering Confucian literature, which likely precluded them from learning the latest information, technology, and scientific discoveries from the West. Squicciarini and Voigtländer (2015) showed that such “upper-tail knowledge” increased productivity and innovation during the Industrial Revolution. Nevertheless, developing the work ethic necessary to attain elite status in society may have conferred advantages on the examinee’s descendants.

Third, to what degree the culture of learning affected people is unclear. After the abolition of the Chinese imperial examination system, in which the quota on the number of people eligible to take the imperial examination was high, more Chinese went abroad to study (Bai, 2019). But it is unclear what effect this had on the majority of people who were not able to go abroad.

Fourth, a unique historical fact allows us to identify a culture of learning from the transfer of knowledge via written language. The contemporary Roman-based Vietnamese script[3] (*Quốc ngữ* in Vietnamese) completely replaced Chinese characters as the official written language in the 1910s. Valencia (2019) identified a channel of knowledge transfer via written language when examining Jesuit missions in 1609 in South America. This channel of transmission can also explain the findings of Chen et al. (2020), who showed that the population density of Chinese imperial elites (1368–1905) was associated with more years of schooling in 278 present-day prefectural residents. It is therefore possible that knowledge of logograph-based Chinese characters and literature boosted the next generation’s education via kinship.

Fifth, as shown in Figure 1 in Alesina and Giuliano (2015), Vietnam’s cultural values differ from those of China. Compared with China, Vietnam places greater importance on family ties and work ethic than luck and connections[4]. If these cultural values are persistent, we might expect a stronger intergenerational transfer of human capital and work ethic. However, the removal of Chinese characters from written Vietnamese affords us the

[3] The Vietnamese alphabet was invented by Alexandre de Rhodes, an Avignonese Jesuit missionary, in 1651 and is derived from the Portuguese alphabet.
[4] Respondents had two choices: a. “In the long-run, hard work usually brings a better life”; and b. “Hard work does not generally bring success; it is more a matter of luck and connections” (Alesina and Giuliano, 2015).

opportunity to examine the main channel through which the learning culture is shared. In addition, the negative correlation between family ties and individualism (Alesina and Giuliano, 2015) suggests that stronger family ties lead to slower national development (Gorodnichenko and Roland, 2017) and perhaps even weaker development of human capital. Moreover, a stronger work ethic may indicate that the imperial examinations were relatively fair, and that the success of people today is more likely due to their own hard work.

Our findings indicate a persistent relationship between the success of imperial elites and the quantity and quality of present-day education. More specifically, we find that the number and density of imperial elites in the past in a given district is associated with a higher literacy rate, school attendance rate in compulsory primary school and twelfth grade, more average years of schooling, and a higher probability of taking the National Entrance Examinations to University (NEEU) in the present day. In contrast, the correlations were negative when the outcomes were rates for primary school dropouts and school non-enrollment. From the analysis at the individual level, we find that the number and density of imperial elites in the district led to higher age-standardized test scores (z-scores) in the NEEU regardless of the test subjects. Higher z-scores might also be a reason for a higher school attendance rate among the district population aged 18–24 years. We propose several hypotheses for this persistent legacy, such as the independence of villages from central power, heavy reliance on academic decree payrolls, especially for university degrees, and favoritism toward districts in which a bureaucrat or politician was born. These might be among the important factors contributing to the culture of learning and determination to this day.

Our paper contributes to the literature in several ways. First, to our knowledge, this paper is the first to examine the connection between a culture of learning and a qualitive measurement of present-day education. Second, we proved the existence of a learning culture by examining various educational outcomes: literacy rate, school attendance rate, school dropout rate, and school non-enrollment rate. Third, we contribute a novel IV calculating the average distance from each district to the corresponding capital where the national-level imperial examination was administered.

The remainder of the paper is organized as follows. We introduce the data sources and how we integrate the data for analysis in Section 2. Section 3 presents our identification strategy, the construction of the IV, and the methods. We report the results and discuss possible mechanisms behind the results in Section 4. Finally, Section 5 presents the conclusion and notes our research limitations.

2. Data

Our analyses are based on several important historical and contemporary data sources. Combining these with geographical data, we laid the historical data on the contemporary data by using identical concordance district names to construct a dataset of 687 identical districts.

2.1 Historical data

We extracted information from Ngo (2006), who collated the longest known list of imperial elites who passed the national level of the 1075–1919 imperial examinations (see Appendix 1 for description) from many ancient sources (mainly the stela stones and the imperial examination records [*Đăng Khoa Lục* in Vietnamese]) and furthermore noted the home of each imperial elite and the corresponding district as it is known today.

We obtained a list of 2,888 names[5] and home districts of imperial elites from the more than 180 recorded imperial examinations in Vietnam. However, we could only obtain up-to-date concordance geographical locations at the district level for 2,844 imperial elites (covering 198 districts[6]). We counted the number of imperial elites in the list for each district to create our main variable, $elite\ number,$ as well as the density of the imperial elites per square kilometer, $elite\ density$. The number of imperial elites per time period is shown in Appendix 2 and Graph 1, and their distribution by geography is shown in Appendix 5.

[Insert Graph 1 here]

2.2 Contemporary data

2.2.1 Quantity of education

We relied on the 2009 Vietnamese Population and Housing Census (100% or 86.9 million observations—hereafter, "population census") conducted by the General Statistics Office of Vietnam (GSO). The population census recorded whether individuals could read and write. We investigated the variable $literacy\ rate$ among those aged ≥7 years in the district population. The census also included a question on school attendance status (as of April 1,

[5] Ngo (2006) originally listed 2,894 names of imperial elites. However, we found a few duplications because some test takers took the test several times to improve their score and obtain the highest rank.

[6] In the 2009 population census, Vietnamese administrative divisions comprised 63 provinces (Level 1) and 689 districts (Level 2).

2009, which is within the school calendar), which measured our second variable, $attendance\ rate$. The census asked if the individual had ever attended school and whether, at the time of the census, they were still attending school. This information helped us to create a third variable for school dropout, $dropout\ rate$. If the individual had completed their education, the information on the highest grade or educational level attained by the individual was recorded. This measured our converted *years of schooling* for each individual as the fourth variable, for which we considered the age cohorts (18+, 22+, and 25+) most likely to have completed their education.

Furthermore, we calculated several outcomes by age cohort at the district level. If the individual had never attended school, we had a corresponding dummy variable, named $never\ enroll$, and calculated the $never\ enroll$ rate by dividing the total number of those who had never attended school in any age cohort by the district population of the same age cohort. We considered $attendance\ rate$, $dropout\ rate$, and the $never\ enroll$ rate at the district level for compulsory education (primary school) in the 7–10 age cohort. Primary school is important because everyone has to graduate before pursuing higher education later in life. In addition, we considered school $attendance\ rate$ in 2009 for individuals born in 1991 in connection with the examinations to university in 2009.

2.2.2 Quality of education

To assess the quality of education, we examined the data of university placement test scores in 2009 (the NEEU). Each classification of the NEEU is administered by the Vietnamese Ministry of Education and Training (MoET) at exactly the same time and date, with the same test problems, and using a centralized anonymous scoring system. A census of NEEU 0.98 million test takers recorded the test scores of individuals (Vu, 2019) according to 11 pre-determined classifications. For decades, each classification of the NEEU comprised three pre-determined test subjects. We considered the four main classifications, which represent 96% of all 2009 NEEU test takers: A (Mathematics, Physics, and Chemistry), B (Biology, Mathematics, and Chemistry), C (Literature, History, and Geography), and D (Literature, Mathematics, and Foreign language). The sum of the three scores (the total test score, hereafter) is used to determine university placement. Universities use the total test score in a given classification to place individuals according to a pre-determined (i.e., decided several months before the test dates) number of "seats" for a specific field of study. Each university pre-determined one or several test classifications for placements in its faculties and fields of studies[7].

The NEEU test scores, especially the Mathematics scores, best reflect the human capital and investment in quality of education. This is because test takers are expected to devote their best effort to all three subjects, which contain no pass/fail threshold. Scoring is anonymous and conducted at the national level. The passing score is determined after all the answer sheets have been scored and is subsequently announced nationwide. By then, universities have ranked test takers according to the sum of their three test scores. Placements are made, starting with the highest scores and proceeding downward until all the seats are filled. The entrance examinations for each classification are held only once in a year. In 2009, the examinations for classification A were held July 4–5, whereas those for B, C, and D were held July 9–10. It is thus practically impossible to take the examinations for more than two classifications. Accordingly, the majority of test takers self-select into a single classification as early as the tenth grade to have adequate time to prepare for their test. Not surprisingly, entrance examinations are extremely competitive. In 2009, the ratio of test takers to available seats ranged from 1:2 to 1:23 (Vu, 2019).

Using the home districts declared by the NEEU test takers on their test application forms, we linked the test scores for classifications A, B, C, and D with the home districts of imperial elites. We considered only the test takers who were born in 1991 because they formed the highest number of test takers (73%). Also, those individuals likely took NEEU for the first time in their lives[8]. We used the three subject scores and the total scores as the fourth variable. More precisely, we used the standardized ($z - score$) test score for each classification as an outcome (hereafter, "z-score") at the individual level. The descriptive statistics of each classification are provided in Appendix 4.

The act of taking the NEEU presents a selection issue. We considered this selection in the probability of taking the NEEU after the twelfth grade ($taking\ NEEU\ after\ 12th\ grade$) at the district level by dividing the sum of identical test takers over the aggregated number of individuals who were born in 1991 and recorded as attending school as of April 1, 2009 in the population census.

[7] For example, a university might decide to use test scores in classification A to place students in the faculty of engineering with a specialty in mechanics. The university might also use classification B for another specialty, such as biotechnology.

[8] We chose individuals who graduated from high school in May 2009.

2.3 Geographical information system data as controls

Based on Spolaore and Wacziarg (2013), we used geographical information system (GIS) data from several sources to create control variables at the district level. We used 2009 nighttime light data (Version 4 DMSP-OLS Nighttime Lights Time Series) from the US National Oceanic and Atmospheric Administration to construct average nighttime light intensity as a proxy for economic activity and development. We preferred the oldest geographical data because they best represent the past characteristics of the district. The 1992 Global Land Cover Characterization from the United States Geological Survey Earth Resources Observation and Science Center (USGS EROS) was used to generate the urbanization land ratio and cropland ratio for each district. From the USGS EROS Center's Landsat Imagery in 1996, we generated the average elevation for each district. We also calculated the distance from each present-day district center to the coastline (distance to coast) using the shape file provided by the Database of Global Administrative Areas[9].

Finally, we obtained a dataset on 687 districts and a dataset on individuals and their test scores for each classification with all the information necessary for our analysis (see Appendix 3 and 4).

3. Methods

3.1 Identification strategy

Unobservable factors could potentially result in endogeneity issues. For example, examinees needed several years to prepare for imperial examinations. To be eligible to take the national level examination one first had to pass the provincial examination. Upon passing the national level examination, imperial elites could be awarded positions even as high as top-ranked public official in the imperial government (Ngo, 2006). It should be noted that the number of imperial elites in a given area was not random because public schools were not present in every district (Ngo, 2006); thus, lower-level administrations relied mainly on private schools funded by the local community (Ngo, 2006).

To account for these unobservable factors, we implemented an IV (hereafter, interchangeably "$average\ distance$" or "IV") to represent the average distance from each

[9] Available at: www.gadm.org

district to each imperial examination test venue's district for all imperial examinations during 1075–1919. For each imperial examination year, we calculate the arc distance between the present-day center of each district to the district where the imperial examinations were actually held[10]. Based on Tran (1920), we decided the timeline of each dynasty and the corresponding capital, combining the information from Ngo (2006) to determine the imperial examination test venues, which were always where the concurrent emperor was living. This is because the emperor made the final decision on who passed the imperial examination and also interviewed the top-ranked candidates (Ngo, 2006). The distance, in kilometers, is a proxy for the cost of attending each imperial examination. We took the average of these distances for each district to form the IV.

We argue that traveling this average distance was both important and substantially costly to the majority of people. This is because the territory of Vietnam had an S-shape, and was separated by high mountains and rivers between the north and the south. In addition, imperial examination test takers were required to arrive at the test venue several months in advance for preparation (Ngo, 2006).

Our calculated IV is not directly related to the current cost of education for individuals today for several important reasons. First, the capital of Vietnam has changed several times throughout history, which was an exogenous factor for test takers. In addition, Vietnam has experienced multiple national divisions, and has had two or even three "capitals" concurrently. The emperor of each region held their own respective imperial examinations (Ngo, 2006). Ordinary citizens could move freely to take the tests under different emperors which were even in the same year (Ngo, 2006).

Second, since 1075, the territory has almost doubled in size (see Appendix 5). Individuals living in South Vietnam today are not necessarily at a disadvantage. For example, the distance to imperial test venues conferred no advantage on individuals in Hue and Thanh Hoa compared with individuals in the south, especially those in the relatively better developed areas, such as Ho Chi Minh City. In 2009, Ho Chi Minh City had a population of 7.1 million people, 1.77 times higher than that of Hanoi, and was ranked number one in nighttime light

[10] We acknowledge the difference between contemporary district boundaries and the corresponding historical ones in each dynasty. However, we assume contemporary district boundaries as proximate common areas to the village from where the imperial elite originated. Each district has an area of 422 km^2 on average, or a circle with a 12km radius.

intensity[11]. Thanh Hoa and Thua Thien Hue province are far lower in the ranking list compared with Hanoi (ranked 4th with a population of 7.5 million). Thanh Hoa's (Thua Thien Hue's)[12] nighttime light intensity are ranked 26th (41th) and are approximately 0.35 (0.26) times that of Hanoi. The territory gradually expanded southward at different speeds[13], which was also exogenous to the residents of any time. We considered the history of Vietnamese territory expansion when constructing the IV. We set the highest cost for the district to be located in the "future territory" as of the year of the imperial examination. Specifically, we let the distance be 2,272 km[14] in a given "future territory" district for the imperial examination year.

Third, the coefficients of $average\ distance$ to the imperial examination test venue were negative and were statistically significant in explaining the number and density of imperial elites by district, regardless of the fixed effects at the province level (see columns (1)–(8) in Table 1) and controls at the district level.

[Insert Table 1 here]

Where provincial fixed effects were present, the coefficients showed the difference in the cost of travelling to the imperial examination venues among districts within a given province.

This contrasts with the correlations between the IV and the density of present-day primary schools, secondary schools, high schools, colleges, universities, hospitals, and health stations (more precisely, the numbers per 2009 district population)[15].

[Insert Table 2 here]

As per Table 2, the average distance to imperial examination venues had no correlation with the variables for present-day density of educational employees, secondary schools, high schools, colleges, hospitals, and clinics when provincial fixed effects were in present. This suggests that the average distance cannot explain the within-province differences. Contemporary density of general schools by contemporary population was least likely determined by the average distance to the imperial examination test venues. This also suggests that if the average distance has causal effects on contemporary educational achievements, the

[11] We used nighttime light data for 2009 from the National Oceanic and Atmospheric Administration for this comparison.
[12] In 2009, they also had respective populations of only 3.4 and 1 million compared with today's 86 million.
[13] See Dell et al. (2018) for some of specific historical events related to the territorial expansion.
[14] The distance of 2,272 km is how far an individual has to travel on foot from Mong Cai (the farthest point in northeastern Vietnam) to Ca Mau (the farthest point in southeastern Vietnam) today. We also tried other values separately for robustness checks.
[15] These variables were from the 2007 Establishment Census by the GSO. We did the same between the IV and number of contemporary employees in the education and health care sectors.

channel should be via factors other than the contemporary density of local school (and health) endowments. In addition, although contemporary institutions may have helped districts with a higher density of schools (Acemoglu et al., 2014), this was least likely to be related to imperial elites.

In addition, the IV also aligned with suggestions from Angrist and Krueger (2001). The IV had no relationship with the abilities of imperial elites that might have been transferred to following generations via kinship. In the intended estimations with educational outcomes, the abilities of subsequent generations potentially remain in the error terms.

3.2 Methods

We first estimated the aforementioned variables in an ordinary least squares model using the reduced form equation for each unit as a district:

(1) $Outcome = \beta_1.nelite + \beta_2.control + \varepsilon,$

where $nelite$ is the number of imperial elites who claim the district as home. We also used the density of imperial elites per square kilometer in the present-day district area as an alternative. In addition to the control variables mentioned in Section 2.3, the ratio of the Kinh ethnic group ($Kinh\ rate$) in the present-day district population[16] was added to form the $control$. We also considered a probability weight for each district, formed by dividing the 2009 district population by the national population.

We converted (1) into (3) below using the IV, two-stage least squares, and generalized method of moments estimation with robust standard errors. Specifically, in the first stage, we estimated

(2) $nelites = \alpha_1.average\ distance + \alpha_2.C + \epsilon.$

In the second stage, we estimated a model similar to (1) as

(3) $Outcome = \gamma_1.\widehat{nelite} + \gamma_2.C + \omega,$

where C is the set of control variables. C in Equations (2) and (3) used the same controls as in Equation (1). In addition, we also used Equations (2) and (3) for estimating the other outcomes, such as the probability of attending school among those born in 1991, and the probability of

[16] We also had a specification without any control variables, as shown in Online Appendices 1, 3, 6, and 8.

taking the 2009 NEEU among those born in 1991 and also having completed the twelfth grade by 2009.

Using the same procedure as in (2) and (3) but at the individual level, we regressed the z-scores of the 2009 NEEU test takers on the number of imperial elites in the corresponding districts. In addition, C also included a gender dummy and a dummy indicating whether or not the individual was ineligible for some contemporary educational policy privileges (1 if ineligible, and 0 otherwise). Those eligible for educational policy privileges were children of soldiers wounded or killed in combat and other national heroes (according to Vietnamese government regulations).

We report estimations where the number of imperial elites according to district was available as the main results. We used the density of imperial elites as an alternative (see Online Appendix 2, 3, 5 and 6).

4. Results

4.1 Persistent effects on present-day quantity of educational attainment

In general, we found that the number of imperial elites in a given district was associated with a higher present-day quantity of educational attainment, as shown in Table 3. The results are robust even when we used density of imperial elites, as shown in Online Appendices 2 and 3. An additional imperial elite was associated with a 0.2% increase in the literacy rate of the district (see column (11) of Table 3), and translated to an additional 0.12 years of schooling among those aged 22+ and 25+, as evidenced in columns (13)–(14) of Table 3.

[Insert Table 3 here]

4.2 Persistent effects on present-day effort to prolong years of schooling

We found that the number of imperial elites in a given district had a positive and statistically significant relationship with higher school attendance rate in 2009, but a lower number of school dropouts and non-enrollment, as shown in columns (15)–(17). It should be noted that these are the rates for compulsory education. Thus, the results suggested that the presence of imperial elites was associated with greater focus and effort on the first stage of present-day education. In addition, primary school attendance is compulsory and free in

Vietnam; therefore, the dropout and non-enrollment rates are considered pure because they were unlikely to be distorted by the cost of learning.

Furthermore, the school attendance rate was higher among those born in 1991 and those aged 18–24 years in association with $elite\ number$, as shown in columns (18)–(19). This indicates that those residing in the imperial elites' home districts had a higher tendency to pursue more years of schooling, to post-high school education.

4.3 Persistent effects on present-day quality of education

We found that the z-scores of the 2009 NEEU test takers were associated with the number of imperial elites who shared the same home district. An additional imperial elite in the district leads to an increase of up to 0.02 standard deviation on test scores, as shown in Table 4. The results are consistent regardless of classification, test length, and for all subjects except for Biology (statistically insignificant).

[Insert Table 4 here]

In addition, results using Mathematics best reflected the persistent legacy on quality of education and human capital. Mathematics was a compulsory subject for all grades. Nearly all (98%) of the test takers had to take the Mathematics test as part of the NEEU (Vu, 2019). As mentioned earlier, the z-score is considered to reflect the best effort of each test taker because the threshold for selecting students for university placement was exogenous to everyone.

4.4 Robustness checks

First, we conducted all estimations using the density of imperial elites per square kilometer as the main explanatory variable and found that the main results were consistent (see Online Appendices 2, 3, 5 and 6).

Second, we re-ran the estimations without using district GIS control variables. The corresponding results shown in Online Appendices 1, 3, 6, and 8 were similar to those of the original estimations. Of the 26 outcomes, there were only three exceptions: the rate of taking the NEEU in 2009, Foreign Language test scores (statistically insignificant), and Biology test scores (statistically insignificant and negative). These exceptions did not change our main interpretation because the results using the sum of the test scores were consistent across all estimations.

Third, we replicated all estimations using the two additional values for maximum distance applied to the abovementioned "future territory". In these replications, we replaced

2,272 km, which was used in our main reported estimations, with 3,000 or 4,000 km. Our replication results, shown in Online Appendices 9–11, were consistent with the main results.

Fourth, we considered spatial autocorrelation, as suggested by Kelly (2019). Districts might be clustered according to geography rather than administrative boundaries. We applied the procedures by Colella et al. (2019)[17] and tried different assumptions for the distance beyond the observations belonging to the same cluster, including 25, 50, 100, 200, and 300 km. We repeated this with the specifications outlined in Table 3, and, once again, using the density of imperial elites. The standard errors changed slightly; however, the results of the new estimations were consistent with the main results for the IV second stage shown in Table 3 (except for the case mentioned in the previous section).

4.5 Possible channels and reasons for persistent historical legacy

The Vietnamese of today have little incentive to study past knowledge on Confucianism. Imperial examinations and Chinese characters are no longer part of Vietnamese life. To our knowledge, none of the content from Confucian textbooks has been transferred to today's textbooks or influenced today's university entrance examinations or technologies. Therefore, even the forced transfer of Confucian content and knowledge via social interaction would confer no educational advantages on those living in imperial elites' districts today.

Instead, the culture of learning and determination would persist and likely be transferred via informal institutions. This is because present-day formal institutions are based on the communist ideology. Vietnamese imperial institutions were abolished in 1945. Meanwhile, the culture of learning and determination was transferred at the local (i.e., village) level via informal institutions for centuries. Nguyen (2005) examined the village institution via records kept in the village's common house (*Đình làng* in Vietnamese). Nguyen (2005) noted that villagers donated land for the building of schools in 1767. They even had public rice fields dedicated for educational purposes. The money earned from renting out the rice fields would be spent on private schools for everyone in the village. The money was also used to hire good teachers and pay them bonuses. Villagers also designated person to manage the village library. Villagers bought books (for learning and exam preparation) and shared them with learners. Nguyen (2005) found village rules stating that the person in charge of education was required to place Confucian books in direct sunlight on sunny days to help prolong the lives of the books. Nguyen (2005) also found records of a village program encouraging villagers to learn, giving

[17] Specifically, we used Stata command "acreg" (Colella et al., 2019), which supports the IV approach as well.

learners an exemption from forced labor such as public service and offering a bonus when learners passed their examinations. The program also encouraged learners to attain the highest rank in the national level of the imperial examinations. The names of those who succeeded would be carved in the stelae or documented in records kept in the village common house. This is still considered a source of family pride today, maintained from generation to generation.

Formal institutions allowed village culture and its regulations to prosper, especially because administrative control at the village level has always been free from central government control (Dell et al., 2018). Present-day government rules (see Government Decree No 92/2009/ND-CP dated November 2009) allow only a few administrative positions at the commune level to receive a salary from the government budget. Moreover, the mandated Government Decree No 34/2019/ND-CP dated April 2019, for the first time, recognized fewer than three positions at the village level, listed them as non-specialized positions, and let them be eligible for a modest subsidy from the government budget. This suggests that village officials are elected mainly based on the will of local voters and the available funding.

There are several potential economic reasons that could explain persistent legacies. First, academic degrees, especially those from institutions of higher learning, play an important role in the wage scale (Vu and Yamada, 2018, 2020). Therefore, pursuing more years of schooling should lead to a higher income, and consequently lift an individual out of poverty. Meanwhile, entrance examinations to university are competitive. Demand far exceeds the number of available seats at universities and only 4%–50% of entrance examination test takers are admitted to universities (Vu, 2019). Tuition fees at universities are heavily subsidized by the government because the majority of universities are public institutions. Second, home favoritism is common among public officials (Do et al., 2017) and top-ranked politicians (Vu and Yamada, 2017). Once individuals had reached these high positions, they would divert resources to their hometown districts.

Kinship and family ties are also important (Wantchekon et al., 2013) and provide a straightforward channel for the legacy. Naturally, some imperial elites were related. Unfortunately, we were not able to further investigate the kinship channel in the manner of Chen et al. (2020) with respect to the Chinese imperial examination system. Further complicating the issue is how common some Vietnamese family names are. For example, the family name Nguyen accounted for nearly a third of both imperial elites and 2009 NEEU test takers. In addition, a given family name written in the Vietnamese alphabet today might be associated with several different names in the past written with Chinese characters that had the same pronunciation. Finally, Vietnamese middle names and the lack of consistency in naming rules made all our attempts at identifying kinship futile.

5. Conclusions

In this paper, we examined the relationship between the number (or density) of imperial elites and the quantity and quality of present-day education in Vietnam. We found that the association is positively statistically significant for educational variables such as average years of schooling (among those aged 18+, 22+, and 25+ years), literacy rate (among those aged ≥7 years), and school attendance rate at the district level. NEEU test scores were positively proportional to the number and density of imperial elites. The results were obtained using the novel IV, the average distance from each district to the former capital where the imperial examinations were held.

This study had several limitations. First, the data we collected on imperial elites do not include every imperial elite that ever existed in Vietnamese history. Some names are missing because records were destroyed, ruined, or lost. The names and homes of imperial elites in the dataset were taken from available and up-to-date documents. Therefore, our results might be subject to bias. Second, assuming a linear travel time cost according to distance, our constructed average distance for IV includes measurement errors. Third, the present-day district boundaries might not be identical to those in the past. Similarly, the culture of learning would be rooted at the village level rather than the district level. Therefore, if information were to be identified at the village level, the effects may be even higher than those estimated. Fourth, we had little confidence asserting a difference in the rate of t NEEU test-taking between those living in the imperial elites' home districts and those living in other districts. This is because we found that the results for this specific outcome were not consistent or significant across specifications and spatial correlation specifics. Fifth, we did not have the information necessary to account for internal migration among imperial elites, although we acknowledged the movement among present-day Vietnamese, which was accounted for 6.6% of the 15% population sample in 2009. Sixth, we did not have sufficiently detailed data about several historical events that occurred between 1919 and 2009. For example, we lacked detailed information on the land reforms implemented between 1953 and 1956 in North Vietnam. Therefore, we were unable to test whether the culture of learning would have been interrupted accordingly, as examined in China by Chen et al. (2020) and by Meng and Grogory (2002).

Similarly, we were unable to consider the re-unification event in 1975 and reforms specific to South Vietnam in our analysis. In addition, we did not consider whether the impact of French colonization might have altered the estimated results, as previous literature (Michalopoulos and Papaioannou, 2020) on Africa did. Given these limitations, we speculate that the legacy of imperial examinations may have been even more profound than our estimates suggest.

References

Acemoglu, D., Gallego, F. A., & Robinson, J. A. (2014). Institutions, human capital, and development. *Annual Review of Economics*, *6*(1), 875–912. https://doi.org/10.1146/annurev-economics-080213-041119

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, *91*(5), 1369–1401. https://doi.org/10.1257/aer.91.5.1369

Alesina, A., & Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, *53*(4), 898–944. https://doi.org/10.1257/jel.53.4.898

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(4), 69–85. https://doi.org/10.1257/jep.15.4.69

Bai, Y. (2019). Farewell to Confucianism: The modernizing effect of dismantling China's imperial examination system. *Journal of Development Economics*, *141*(September), 102382. https://doi.org/10.1016/j.jdeveco.2019.102382

Becker, S. O., Boeckh, K., Hainz, C., & Woessmann, L. (2016). The empire is dead, Long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *The Economic Journal*, *126*(590), 40–74. https://doi.org/10.1111/ecoj.12220

Dell, M., Lane, N., & Querubin, P. (2018). The historical state, local collective action, and economic development in Vietnam. *Econometrica*, *86*(6), 2083–2121. https://doi.org/10.3982/ECTA15122

Do, Q., Nguyen, K.-T., & Tran, A. N. (2017). One Mandarin Benefits the Whole Clan: Hometown Favoritism in an Authoritarian Regime. *American Economic Journal: Applied Economics*, *9*(4), 1–29. https://doi.org/10.1257/app.20130472

Chen, T., Kung, J. K.-S., & Ma, C. (2020). Long live Keju! The persistent effects of China's civil examination system. *The Economic Journal*, (June). https://doi.org/10.1093/ej/ueaa043

Colella, F., Lalive, R., Sakalli, S. O., & Thoenig, M. (2019). Inference with Arbitrary Clustering. *IZA Discussion Paper Series*, (12584), 1–15.

Gorodnichenko, Y., & Roland, G. (2017). Culture, institutions, and the wealth of nations. *The Review of Economics and Statistics*, *99*(3), 402–416. https://doi.org/10.1162/REST_a_00599

Kelly, M. (2019). The standard errors of persistence. *UCD Centre for Economic Research Working Paper Series*, No. WP 19/13.

Meng, X., & Gregory, R. G. (2002). The impact of interrupted education on subsequent educational attainment: A cost of the Chinese Cultural Revolution. *Economic Development and Cultural Change*, *50*(4), 935–959. https://doi.org/10.1086/342761

Michalopoulos, S., & Papaioannou, E. (2020). Historical legacies and African development. *Journal of Economic Literature*, *58*(1), 53–128. https://doi.org/10.1257/jel.20181447

Ngo (2006). List of Vietnamese imperial elites 1075-1919 ("Các nhà khoa bảng Việt Nam", in Vietnamese). Edited by Ngo Duc Tho, Literature Publishing House: Hanoi.

Nguyen, H. M. (2005). Several documented imperial education at village and commune level from recorded stelae ("Vài tư liệu về tình hình giáo dục Nho học ở cấp làng xã qua tư liệu văn bia" in Vietnamese). Shino-Nom proceedings ("Tạp chí Hán Nôm" in Vietnamese), 4(71).

Nunn, N. (2020). The historical roots of economic development. *Science*, *367*(6485). https://doi.org/10.1126/science.aaz9986

Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature*, *51*(2), 325–369. https://doi.org/10.1257/jel.51.2.325

Squicciarini, M. P., & Voigtländer, N. (2015). Human capital and industrialization: Evidence from the Age of Enlightenment. *The Quarterly Journal of Economics*, *130*(4), 1825–1883. https://doi.org/10.1093/qje/qjv025

Thuan Hoa. (2010). "Ngo Bao Chau on the date to pay tribute to his ancestors for his success" ('Ngô Bảo Châu trong ngày "vinh quy bái tổ"' in Vietnamese). Link: https://vietnamnet.vn/tinnhanh/201009/Ngo-Bao-Chau-ngay-vinh-quy-bai-to-934282/. Last access April 1, 2020.

Tran, T. K. (1920). A brief history of Vietnam (Việt-Nam Sử-Lược, in Vietnamese), Edited by Tran Trong Kim. Literature Publish, 2008: Hanoi, Vietnam.

Valencia C. F. (2019). The mission: Human capital transmission, economic persistence, and culture in South America. *The Quarterly Journal of Economics*, *134*(1), 507–556. https://doi.org/10.1093/qje/qjy024

Vu, T. M. (2019). Effect of weather on human capital in Vietnam. *AGI Working Paper Series, 2019-01*. Link: http://id.nii.ac.jp/1270/00000150/ (Last accessed April 1, 2020).

Vu, T. M., & Yamada, H. (2017). Firms and regional favoritism. *AGI Working Paper Series, 2017-16*. Link: http://id.nii.ac.jp/1270/00000136/ (Last accessed April 1, 2020).

Vu, T. M., & Yamada, H. (2018). Decomposing Vietnamese gender equality in terms of wage distribution. *Pacific Economic Review*, *23*(5), 705–731. https://doi.org/10.1111/1468-0106.12269

Vu, T. M., & Yamada, H. (2020). Convergence of public and private enterprise wages in a transition economy: Evidence from a distributional decomposition in Vietnam, 2002–2014. *Economic Systems*, *44*(1), 100737. https://doi.org/10.1016/j.ecosys.2019.100737

Wantchekon, L., Novta, N., Klasnja, M., & Novta, N. (2013). Education and human capital externalities: Evidence from Colonial Benin. *The Quarterly Journal of Economics*, *130*(2), 52. https://doi.org/10.1093/qje/qjv004

Appendix 1 Brief description of 1075–1919 imperial examinations in Vietnam

Vietnamese emperors used imperial examinations to select state bureaucrats from among the citizenry starting in 1075 (in the Ly Dynasty). Inspired by the Chinese imperial examination system, the Vietnamese examinations were initially based on the teachings (*Tam giáo* in Vietnamese) of Confucianism, Taoism, and Buddhism but from 1232 focused solely on Confucianism and its literature (Ngo, 2006). From 1396, the imperial examinations were split into two tiers (Ngo, 2006), the provincial level (*thi Hương* in Vietnamese) and the national level (*thi Hội* in Vietnamese). Only those who passed the provincial examination were eligible to take the national examination. During the Ho Dynasty (1400–1407), they were even asked to visit the capital to study for the national examination for almost a year (Ngo, 2006).

We counted only those who passed the national level examination. The selection rate for the national examination was as low as 1% of all test takers in 1463 (Ngo, 2006). The national level had one or two rounds of tests, depending on the dynasty. The first level was a written test and the second, which was probably only offered to those with the highest scores on the written test, was an oral examination (*thi Đình* in Vietnamese)[18].

Ngo (2006) reported that the national examinations were evaluated blindly. For example, in 1442, the original answer sheets were hand-copied, sealed, and stored carefully. Two independent evaluators anonymously evaluated the examinations according to their copy of the answer sheet and submitted the scores to the Chief Examinations Executive, who created a short list of those who passed and those who would be interviewed by the emperor. With this assistance of his bureaucrats, the emperor would directly interview the shortlisted candidates and select the top ones (the winners).

The number of winners and runners-up as well as the awarded title names varied over time. From 1247, the highest ranked test taker was deemed the principal graduate (*Trạng Nguyên* in Vietnamese, or 狀元 in Chinese). The special treatment offered by the emperor to the winners and runners-up from 1442 included an emperor banquet, clothes, servants, horses, and permission to perform a special ritual (*vinh quy bái tổ* in Vietnamese) to report their success to their ancestors in their hometown. In addition, the highest-scoring winners' names were carved in a stone stela and stored in the national culture temples, while the names of all those who passed the national level examinations (written form) (including the winners) were recorded in a book entitled *List of Imperial Elites* (*Đăng Khoa Lục* in Vietnamese).

[18] However, sometimes the interviews were not conducted (Ngo, 2006). Those who passed the written form were treated as imperial elites in our analysis.

Appendix 2 Test venues, timelines, and number of recorded imperial elites

Year of exams	Dynasty	Test venue		Raw data	
		Province	District	Times	Elites
1075-1225	Ly	Hanoi	Ba Dinh	4	11
1232-1393	Tran	Hanoi	Ba Dinh	16	49
1400-1405	Ho	Thanh Hoa	Vinh Loc	3	11
1426-1526	Le	Hanoi	Ba Dinh	32	1,008
1529-1592	Mac	Hanoi	Ba Dinh	22	482
1554-1595	Le	Thanh Hoa	Tho Xuan	8	51
1598-1787	Le	Hanoi	Ba Dinh	64	723
1822-1919	Nguyen	Thua Thien Hue	Hue	39	553
Sum				188	2,888

Appendix 3 Descriptive statistics of district data

Variable	Mean	Std. Dev.	Min	Max
Number of elites (*Elite number*)	4.1325	11.2106	0	98
Density of elites (*Elite density*)	0.0547	0.3444	0	7.7807
The IV - $Average\ distance$ to imperial test venue (km)	999.9595	845.5751	23.0431	2127.5340
District population	126,227	87,751	83	701,194
District area (km^2)	435.0914	403.3787	1.542285	2678.948
Urban land rate	0.0228	0.1147	0	1
Cropland rate	0.4698	0.3166	0	1
Elevation mean (m)	216.2590	309.7120	1	1544.59
Distance to coastal line (km)*	86.11	84.77	0.23	434.57
Nightlight intensity	8.3877	14.1880	0	63
Literate rate	0.9202	0.0894	0.4717	1
Kinh ethnic rate	0.6724	0.2968	0.0156	0.9114
Years of schooling				
18+	9.0020	1.4968	6.1028	17.5763
22+	7.6128	1.2696	5.3661	13.4042
25+	7.4610	1.2767	5.1565	13.1569
Attending school rate				
1991-born*	0.5491	0.1760	0.1429	0.9182
18-24	0.1675	0.1349	0.0451	0.7476
*Compulsory education among aged 7-10**				
Attending rate	0.9627	0.0442	0.6077	1
Dropout rate	0.0153	0.0123	0	0.0692
Never-enroll rate	0.0218	0.0360	0	0.3758
NEEU taking rate **	0.4482	0.1451	0	1

Notes:

Number of districts = 687

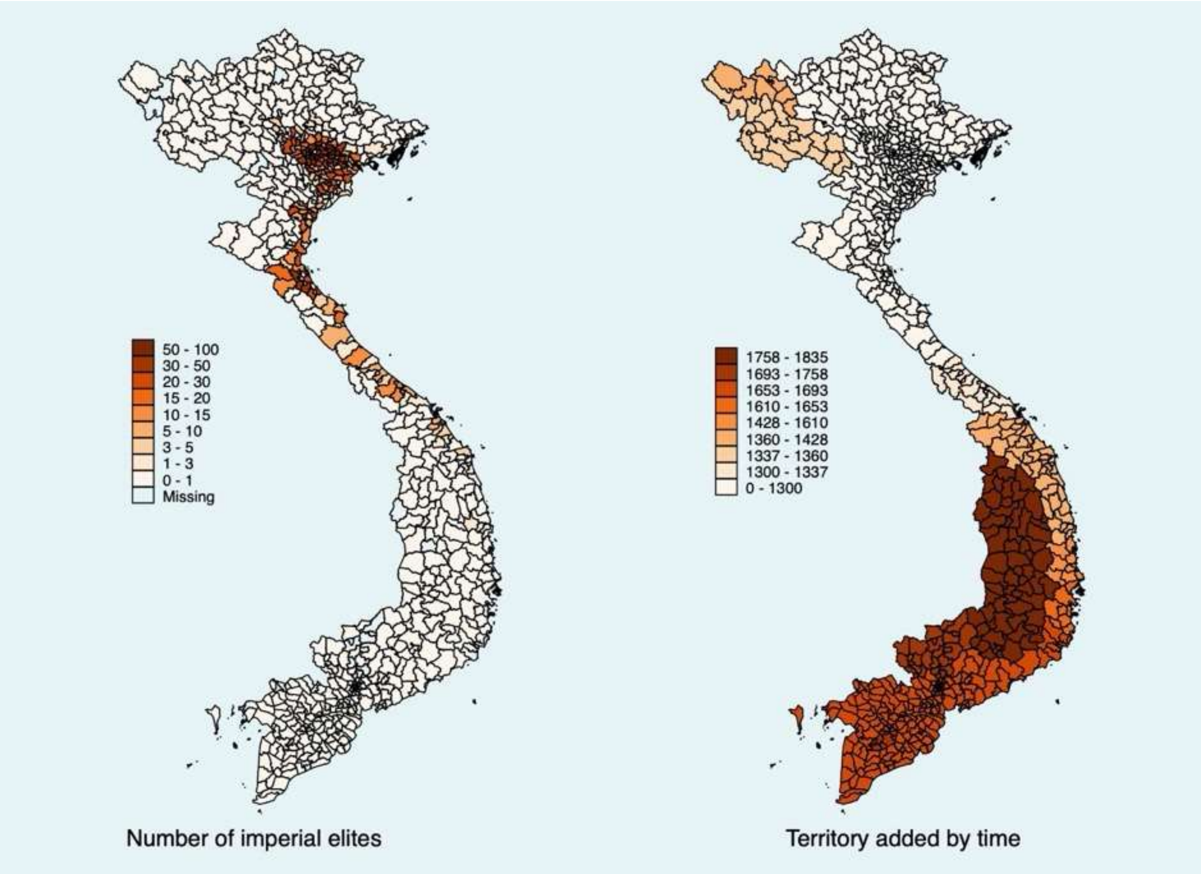
* Number of districts = 686 (1 island district did not have any eligible individuals)

** Number of districts = 681; the rate was among those born in 1991 who finished the twelfth grade

Appendix 4 Descriptive statistics of 2009 NEEU test scores

Variable	Mean	Std. Dev.	Min	Max
A classification				
N= 340,855				
Z-score				
Sum	0.0003	1.0001	–1.3259	2.8196
Physics	0.0003	1.0001	–1.7885	4.2505
Mathematics	0.0002	1.0001	–3.3757	2.6521
Chemistry	0.0003	1.0002	–2.6303	4.6248
Sex (=1, if female)	0.5246	0.4994	0	1
No privilege	0.8998	0.3003	0	1
B classification				
N=178,704				
Z–score				
Sum	0.0004	1.0001	–1.4543	2.7257
Biology	0.0001	1.0000	–3.2974	4.0949
Mathematics	0.0004	1.0002	–1.5672	3.1413
Chemistry	0.0004	1.0002	–2.8163	3.9163
Sex (=1, if female)	0.5377	0.4986	0	1
No privilege	0.9025	0.2966	0	1
C classification				
N=50,721				
Z-score				
Sum	0.0003	1.0000	–1.6403	2.9178
Literature	0.0012	0.9996	–3.4451	3.0550
History	–0.0005	1.0000	–1.4652	3.6285
Geography	0.0002	0.9998	–3.4745	2.9814
Sex (=1, if female)	0.8166	0.3870	0	1
No privilege	0.8532	0.3540	0	1
D classification				
N=111,029				
Z–score				
Sum	0.0004	1.0000	–1.3448	3.2335
Literature	0.0004	1.0000	–3.6740	3.4757
Mathematics	0.0004	1.0001	–1.4702	3.2093
Foreign language	0.0001	1.0001	–2.4647	4.2135
Sex (=1, if female)	0.8083	0.3937	0	1
No privilege	0.9447	0.2285	0	1

Appendix 5 Number of imperial elites and approximate in-land territory expansion timelines



Note:
The map shape corresponds with the 2015 division. Therefore, "missing" cells indicate newly established districts rather than missing data in 2009. Timelines are approximations based on Tran (1919) used for the calculation of the IV only. We do not claim any historical evidence.

Online Appendix 1 Effects on present-day educational achievements (no other district controls)

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Elite number	0.0011***	0.0408***	0.0340***	0.0335***	0.0007***	–0.0003***	–0.0004***	0.0056***	0.0020**	0.0007
	(0.0001)	(0.0092)	(0.0052)	(0.0051)	(0.0001)	(0.0000)	(0.0000)	(0.0006)	(0.0009)	(0.0007)
R-squared	0.047	0.086	0.095	0.094	0.082	0.136	0.042	0.175	0.028	0.006
IV 2nd stage	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Elite number	0.0015***	0.1074***	0.0964***	0.0967***	0.0020***	–0.0012***	–0.0008***	0.0167***	0.0043***	–0.0000
	(0.0003)	(0.0166)	(0.0137)	(0.0138)	(0.0002)	(0.0001)	(0.0001)	(0.0018)	(0.0013)	(0.0009)
F-statistics†	94.45	94.45	94.45	94.45	94.45	94.45	94.45	94.45	94.45	94.46
N districts	687	687	687	687	686	686	686	686	687	681

Notes:
† Kleibergen-Paap Wald rk F statistic for testing H0: Weak identification test.
Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). P-weight was in all estimations.

Online Appendix 2 Effects on present-day educational achievements (using elite density)

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Elite density	0.0058*	0.8425	0.6819*	0.6772*	0.0086*	–0.0060*	–0.0027*	0.0797	0.0523	0.0262***
	(0.0033)	(0.5618)	(0.3891)	(0.3799)	(0.0047)	(0.0032)	(0.0016)	(0.0502)	(0.0418)	(0.0074)
R-squared	0.562	0.575	0.560	0.550	0.304	0.180	0.364	0.251	0.488	0.466
IV 2nd stage	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Elite density	0.1450***	8.7584***	7.7896***	7.8107***	0.1505***	–0.0858***	–0.0661***	1.2195***	0.4296***	0.1186**
	(0.0326)	(1.7258)	(1.5520)	(1.5596)	(0.0319)	(0.0178)	(0.0147)	(0.2513)	(0.0942)	(0.0515)
F-statistics†	23.83	23.83	23.83	23.83	23.83	23.83	23.83	23.83	23.83	23.72
N districts	687	687	687	687	686	686	686	686	687	681

Note: Same as Table 3.

Online Appendix 3 Effects on present-day educational achievements (using elite density but no other controls)

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Elite density	0.0218**	1.6459**	1.3390**	1.3238**	0.0119**	–0.0056**	–0.0064**	0.1172**	0.1254**	0.0451***
	(0.0099)	(0.8230)	(0.5862)	(0.5723)	(0.0055)	(0.0026)	(0.0030)	(0.0557)	(0.0627)	(0.0134)
R-squared	0.018	0.132	0.140	0.138	0.020	0.032	0.010	0.071	0.098	0.020
IV 2nd stage	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Elite density	0.1028***	7.2051***	6.4672***	6.4872***	0.1343***	–0.0825***	–0.0532***	1.1197***	0.2885***	–0.0015
	(0.0282)	(1.4922)	(1.3222)	(1.3286)	(0.0306)	(0.0186)	(0.0132)	(0.2416)	(0.0865)	(0.0609)
F-statistics†	20.31	20.31	20.31	20.31	20.31	20.31	20.31	20.31	20.31	20.31
N districts	687	687	687	687	686	686	686	686	687	681

Note: Same as Table 3.

Online Appendix 4 Effects on present-day educational achievements using elite number and accounting for spatial correlations

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
Cutoff = 25km										
Elite number	0.0023***	0.1399***	0.1244***	0.1247***	0.0024***	–0.0014***	–0.0011***	0.0195***	0.0069***	0.0019*
	(0.0004)	(0.0162)	(0.0147)	(0.0149)	(0.0004)	(0.0002)	(0.0002)	(0.0024)	(0.0013)	(0.0010)
Cutoff = 50km										
Elite number	0.0023***	0.1399***	0.1244***	0.1247***	0.0024***	–0.0014***	–0.0011***	0.0195***	0.0069***	0.0019
	(0.0006)	(0.0218)	(0.0204)	(0.0208)	(0.0005)	(0.0003)	(0.0002)	(0.0033)	(0.0014)	(0.0013)
Cutoff = 100km										
Elite number	0.0023***	0.1399***	0.1244***	0.1247***	0.0024***	–0.0014***	–0.0011***	0.0195***	0.0069***	0.0019
	(0.0006)	(0.0234)	(0.0217)	(0.0220)	(0.0005)	(0.0003)	(0.0003)	(0.0033)	(0.0016)	(0.0016)
Cutoff = 200km										
Elite number	0.0023***	0.1399***	0.1244***	0.1247***	0.0024***	–0.0014***	–0.0011***	0.0195***	0.0069***	0.0019
	(0.0005)	(0.0172)	(0.0151)	(0.0154)	(0.0004)	(0.0002)	(0.0002)	(0.0021)	(0.0014)	(0.0016)
Cutoff = 300km										
Elite number	0.0069***	0.0023***	0.1399***	0.1244***	0.1247***	–0.0014***	–0.0011***	0.0195***	0.0069***	0.0019
	(0.0014)	(0.0005)	(0.0147)	(0.0130)	(0.0133)	(0.0002)	(0.0002)	(0.0019)	(0.0012)	(0.0015)
N districts	687	687	687	687	686	686	686	686	687	681

Note:
All estimations included the Kinh ethnic rate, nighttime light intensity, population density in 2009, urban land ratio, and cropland ratio in 1992, mean elevation in 1996, and distance to the coastline as controls, and used p-weight.
Colella et al. (2019) specifies cutoff as the lower-limit distance. If two districts were located farther than the lower limit, the correlation between error terms of the two is assumed to be zero. We used the "bartlett" option in the Stata command "acreg" (see Colella et al., 2019 for an explanation).

Online Appendix 5 Effects on present-day educational achievements using elite density and accounting for spatial correlations

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
Cutoff = 25km										
Elite density	0.1450**	8.7584***	7.7896***	7.8107***	0.1505**	–0.0858**	–0.0661**	1.2195***	0.4296***	0.1186
	(0.0620)	(2.9796)	(2.7801)	(2.8099)	(0.0609)	(0.0336)	(0.0280)	(0.4634)	(0.1203)	(0.0770)
Cutoff = 50km										
Elite density	0.1450**	8.7584***	7.7896**	7.8107**	0.1505**	–0.0858**	–0.0661**	1.2195**	0.4296***	0.1186
	(0.0692)	(3.2578)	(3.0549)	(3.0926)	(0.0664)	(0.0366)	(0.0307)	(0.4961)	(0.1266)	(0.1011)
Cutoff = 100km										
Elite density	0.1450**	8.7584***	7.7896***	7.8107***	0.1505**	–0.0858**	–0.0661**	1.2195***	0.4296***	0.1186
	(0.0679)	(3.1351)	(2.9320)	(2.9670)	(0.0647)	(0.0354)	(0.0304)	(0.4649)	(0.1213)	(0.1145)
Cutoff = 200km										
Elite density	0.1450***	8.7584***	7.7896***	7.8107***	0.1505***	–0.0858***	–0.0661**	1.2195***	0.4296***	0.1186
	(0.0558)	(2.6780)	(2.4415)	(2.4647)	(0.0554)	(0.0298)	(0.0267)	(0.3779)	(0.1073)	(0.1111)
Cutoff = 300km										
Elite density	0.1450***	8.7584***	7.7896***	7.8107***	0.1505***	–0.0858***	–0.0661***	1.2195***	0.4296***	0.1186
	(0.0490)	(2.3825)	(2.1478)	(2.1663)	(0.0487)	(0.0260)	(0.0236)	(0.3328)	(0.1030)	(0.1010)
N districts	687	687	687	687	686	686	686	686	687	681

Note:
Same as Online Appendix 4.

Online Appendix 6 Effects on 2009 NEEU z-scores (without other district controls)

Classification	**A**				**B**			
Variables	Sum	Physics	Mathematics	Chemistry	Sum	Biology	Mathematics	Chemistry
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elite number	0.0063***	0.0073***	0.0039***	0.0055***	0.0081***	0.0012**	0.0101***	0.0069***
	(0.0007)	(0.0009)	(0.0006)	(0.0006)	(0.0010)	(0.0006)	(0.0012)	(0.0009)
R-squared	0.012	0.011	0.011	0.010	0.024	0.009	0.030	0.015
IV 2nd stage	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Elite number	0.0167***	0.0200***	0.0088***	0.0151***	0.0188***	-0.0022*	0.0262***	0.0157***
	(0.0020)	(0.0023)	(0.0016)	(0.0017)	(0.0022)	(0.0012)	(0.0027)	(0.0020)
F–statistics†	135.79	135.79	135.79	135.79	132.47	132.47	132.47	132.47
N test takers	340,855	340,855	340,855	340,855	178,704	178,704	178,704	178,704
Classification	**C**				**D**			
Variables	Sum	Literature	History	Geography	Sum	Literature	Mathematics	Foreign language
OLS	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Elite number	0.0062***	0.0100***	0.0031***	0.0026***	0.0058***	0.0118***	0.0039***	–0.0013
	(0.0008)	(0.0010)	(0.0007)	(0.0008)	(0.0016)	(0.0015)	(0.0013)	(0.0013)
R-squared	0.024	0.063	0.009	0.014	0.007	0.074	0.014	0.004
IV 2nd stage	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Elite number	0.0222***	0.0406***	0.0086***	0.0079***	0.0248***	0.0412***	0.0176***	0.0009
	(0.0024)	(0.0037)	(0.0018)	(0.0017)	(0.0045)	(0.0046)	(0.0034)	(0.0036)
F-statistics†	133.29	133.29	133.29	133.29	106.03	106.03	106.03	106.03
N test takers	50,721	50,721	50,721	50,721	111,029	111,029	111,029	111,029

Notes:
† Kleibergen-Paap Wald rk F statistic for testing H0: Weak identification test.
Robust district-clustered standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
All estimations included sex and $no\ privilege$ as controls.

Online Appendix 7 Effects on 2009 NEEU z-scores (using elite density)

Classification	**A**				**B**			
Variables	Sum	Physics	Mathematics	Chemistry	Sum	Biology	Mathematics	Chemistry
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elite density	0.0570	0.0733*	0.0215	0.0557	0.1162*	–0.0159*	0.1498*	0.1166*
	(0.0358)	(0.0401)	(0.0185)	(0.0365)	(0.0678)	(0.0090)	(0.0812)	(0.0658)
R-squared	0.040	0.037	0.035	0.023	0.064	0.024	0.065	0.050
IV 2nd stage	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Elite density	1.0337***	1.1692***	0.5851***	0.9668***	1.8171***	–0.1139	2.4069***	1.6247***
	(0.3129)	(0.3510)	(0.1848)	(0.2927)	(0.3980)	(0.0846)	(0.5229)	(0.3562)
F–statistics†	11.35	11.35	11.35	11.35	21.99	21.99	21.99	21.99
N test takers	340,855	340,855	340,855	340,855	178,704	178,704	178,704	178,704
Classification	**C**				**D**			
Variables	Sum	Literature	History	Geography	Sum	Literature	Mathematics	Foreign language
OLS	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Elite density	0.1171***	0.2175***	0.0460**	0.0384	0.0798**	0.1155***	0.0597*	0.0132*
	(0.0380)	(0.0811)	(0.0203)	(0.0264)	(0.0353)	(0.0442)	(0.0313)	(0.0075)
R-squared	0.038	0.090	0.019	0.020	0.079	0.092	0.060	0.084
IV 2nd stage	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Elite density	2.1209***	4.1008***	0.4318**	1.0493***	0.8585***	1.2571***	0.6565***	0.1086**
	(0.3869)	(0.6890)	(0.2046)	(0.2600)	(0.3132)	(0.4584)	(0.2427)	(0.0543)
F-statistics†	37.37	37.37	37.37	37.37	7.39	7.39	7.39	7.39
N test takers	50,721	50,721	50,721	50,721	111,029	111,029	111,029	111,029

Note:
Same as Table 4.

Online Appendix 8 Effects on 2009 NEEU z-scores (using elite density but no other controls)

Classification	**A**				**B**			
Variables	Sum	Physics	Mathematics	Chemistry	Sum	Biology	Mathematics	Chemistry
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elite density	0.1339**	0.1579**	0.0886**	0.1020*	0.2738**	0.0829**	0.3049**	0.2477**
	(0.0658)	(0.0725)	(0.0440)	(0.0558)	(0.1256)	(0.0356)	(0.1418)	(0.1137)
R-squared	0.009	0.007	0.010	0.007	0.020	0.009	0.020	0.013
IV 2nd stage	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Elite density	1.0155***	1.2160***	0.5329***	0.9189***	1.7468***	–0.2080*	2.4382***	1.4623***
	(0.2743)	(0.3236)	(0.1567)	(0.2499)	(0.3125)	(0.1182)	(0.4228)	(0.2659)
F–statistics†	13.21	13.21	13.21	13.21	34.09	34.09	34.09	34.09
N test takers	340,855	340,855	340,855	340,855	178,704	178,704	178,704	178,704

Classification	**C**				**D**			
Variables	Sum	Literature	History	Geography	Sum	Literature	Mathematics	Foreign language
OLS	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Elite density	0.2142**	0.3827**	0.1225***	0.0364	0.1773***	0.1608**	0.1311**	0.1166***
	(0.0872)	(0.1609)	(0.0396)	(0.0322)	(0.0680)	(0.0640)	(0.0544)	(0.0377)
R-squared	0.020	0.055	0.009	0.012	0.022	0.067	0.023	0.013
IV 2nd stage	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Elite density	2.3319***	4.2721***	0.9066***	0.8315***	0.7794***	1.2921***	0.5535***	0.0272
	(0.3561)	(0.6052)	(0.2120)	(0.1992)	(0.2807)	(0.4886)	(0.2021)	(0.1082)
F-statistics†	49.86	49.86	49.86	49.86	6.47	6.47	6.47	6.47
N test takers	50,721	50,721	50,721	50,721	111,029	111,029	111,029	111,029

Note:
Same as Online Appendix 6.

Online Appendix 9 Replicated main estimations in Table 1 with different values of maximum (max) distance to construct the IV

Elite	Number	Number	Number	Number	Density	Density	Density	Density
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Max=3000km								
Average distance	–0.0057***	–0.0591**	–0.0057***	–0.0526**	–0.0001***	–0.0030***	–0.0001***	–0.0021***
	(0.0006)	(0.0255)	(0.0005)	(0.0229)	(0.0000)	(0.0010)	(0.0000)	(0.0006)
R-squared	0.233	0.584	0.364	0.620	0.054	0.242	0.201	0.365
Max=4000km								
Average distance	–0.0043***	–0.0591**	–0.0043***	–0.0526**	–0.0001***	–0.0030***	–0.0001***	–0.0021***
	(0.0004)	(0.0255)	(0.0004)	(0.0230)	(0.0000)	(0.0010)	(0.0000)	(0.0006)
R-squared	0.228	0.584	0.360	0.621	0.052	0.242	0.199	0.365
Province FE		Yes		Yes		Yes		Yes
N districts	687	687	687	687	687	687	687	687

Note:
Same as Table 1.

Online Appendix 10 Replicated main IV 2nd stage estimations in Table 3 with different values of maximum (max) distance to construct the IV

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Max=3000km										
Elite number	0.0023***	0.1407***	0.1248***	0.1251***	0.0024***	–0.0014***	–0.0010***	0.0196***	0.0070***	0.0019**
	(0.0003)	(0.0134)	(0.0121)	(0.0123)	(0.0002)	(0.0001)	(0.0001)	(0.0018)	(0.0011)	(0.0008)
N districts	687	687	687	687	686	686	686	686	687	681
F-statistics†	118.07	118.07	118.07	118.07	118.07	118.07	118.07	118.07	118.07	117.74
B. Max=4000km										
Elite number	0.0022***	0.1411***	0.1250***	0.1252***	0.0024***	–0.0014***	–0.0010***	0.0197***	0.0071***	0.0018**
	(0.0003)	(0.0134)	(0.0121)	(0.0123)	(0.0002)	(0.0001)	(0.0001)	(0.0018)	(0.0011)	(0.0008)
N districts	687	687	687	687	686	686	686	686	687	681
F-statistics†	117.74	117.74	117.74	117.74	117.73	117.73	117.73	117.73	117.74	117.34

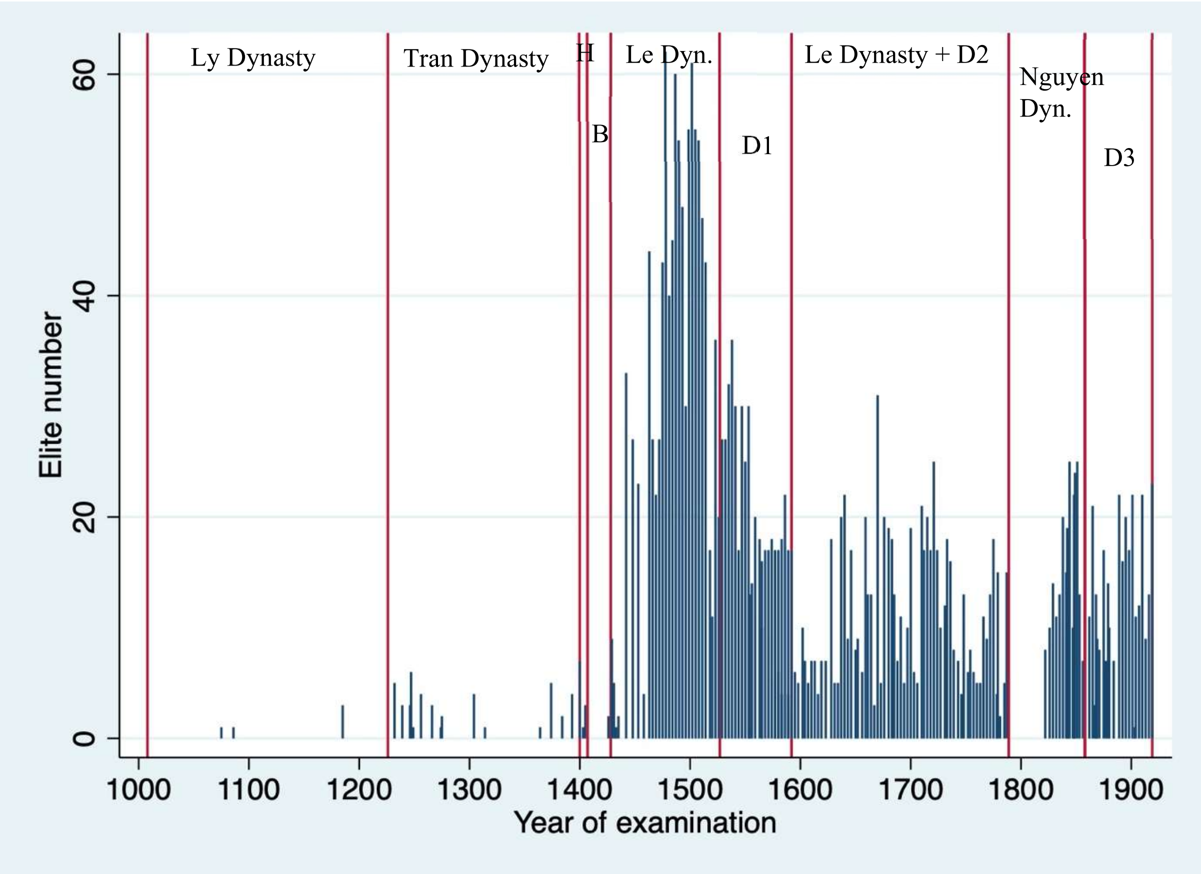
Notes:
Same as Table 3.

Online Appendix 11 Replicated main IV 2nd stage estimations in Table 4 with different values of maximum (max) distance to construct the IV

Classification	**A**				**B**			
Variables	Sum	Physics	Mathematics	Chemistry	Sum	Biology	Mathematics	Chemistry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Max=3000km								
Elite number	0.0217***	0.0245***	0.0122***	0.0204***	0.0236***	–0.0017	0.0313***	0.0212***
	(0.0021)	(0.0023)	(0.0015)	(0.0020)	(0.0023)	(0.0011)	(0.0030)	(0.0022)
F-statistics†	134.05	134.05	134.05	134.05	131.54	131.54	131.54	131.54
Max=4000km								
Elite number	0.0216***	0.0244***	0.0121***	0.0204***	0.0235***	–0.0019*	0.0312***	0.0212***
	(0.0021)	(0.0023)	(0.0015)	(0.0020)	(0.0023)	(0.0011)	(0.0030)	(0.0022)
F–statistics†	133.89	133.89	133.89	133.89	131.42	131.42	131.42	131.42
N test takers	340,855	340,855	340,855	340,855	178,704	178,704	178,704	178,704
Classification	**C**				**D**			
Variables	Sum	Literature	History	Geography	Sum	Literature	Mathematics	Foreign language
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Max=3000km								
Elite number	0.0213***	0.0409***	0.0044**	0.0106***	0.0304***	0.0443***	0.0232***	0.0039**
	(0.0026)	(0.0039)	(0.0020)	(0.0022)	(0.0034)	(0.0048)	(0.0028)	(0.0015)
F-statistics†	132.49	132.49	132.49	132.49	109.41	109.41	109.41	109.41
Max=4000km								
Elite number	0.0212***	0.0407***	0.0044**	0.0107***	0.0303***	0.0443***	0.0232***	0.0040**
	(0.0026)	(0.0039)	(0.0020)	(0.0022)	(0.0034)	(0.0048)	(0.0028)	(0.0015)
F–statistics†	132.93	132.93	132.93	132.93	109.49	109.49	109.49	109.49
N test takers	50,721	50,721	50,721	50,721	111,029	111,029	111,029	111,029

Note:
Same as Table 4.

Graph 1 Number of national winners in the imperial examinations by time



Notes:

- *Ly* Dynasty (1009–1225).
- *Tran* Dynasty (1226–1399.
- H: *Ho* Dynasty (1400–1407).
- B: Chinese domination (1407–1427).
- *Le* Dyn.: *Le* Dynasty (1428–1526).
- D1: Division 1: *Mac* Dynasty (1527–1677) with the capital in Ba Dinh, Hanoi until 1592. *Le* Dynasty during division 1 (1533-1789) with the capital in Tho Xuan, Thanh Hoa from 1533 to 1597.
- *Le* Dynasty + D2: *Le* dynasty (*Le Trung Hung*) (1533–1789) with division 2. *Le* dynasty was under the control of *Trinh* lords (1570–1786), with the capital in Ba Dinh, Hanoi. *Nguyen* Lords (later *Nguyen* Dynasty) (1600–1945) with the capital mainly in Hue. *Mac* dynasty (1527–1677) with the capital in Cao Bang City today from 1593 to 1677. *Tay Son* dynasty (*Nguyen* brothers) (1778–1802) had capitals in Hue and Binh Dinh today, but no imperial examinations.
- *Nguyen* Dyn.: *Nguyen* Dynasty (1600–1945) with the capital in Hue.
- D3: Division 3. French colony (1858–1945). *Nguyen* Dynasty (1600-1945) with the capital in Hue. The last examinations were conducted in 1919.

Table 1 Correlations between imperial elites in the district and average distance

Elite	Number	Number	Number	Number	Density	Density	Density	Density
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
The IV - $Average\ distance$ to imperial test venues	–0.0073***	–0.0591**	–0.0074***	–0.0526**	–0.0001***	–0.0030***	–0.0001***	–0.0021***
	(0.0008)	(0.0255)	(0.0007)	(0.0229)	(0.0000)	(0.0010)	(0.0000)	(0.0006)
2009-Kinh rate			14.9688***	3.8580*			0.2641***	0.0796
			(1.8360)	(2.1970)			(0.0908)	(0.1044)
2009-Night light intensity			0.0051	–0.0827			–0.0002	0.0044**
			(0.0398)	(0.0567)			(0.0020)	(0.0021)
1992-Urban land ratio			–7.1732	–5.2461			–0.0659	–0.1598
			(5.0884)	(5.0553)			(0.5090)	(0.3334)
1992-Cropland ratio			11.3511***	10.0949***			0.0935*	0.0110
			(1.8635)	(2.3611)			(0.0477)	(0.0981)
1996-Elevation			0.0071***	–0.0015			0.0001***	0.0001
			(0.0020)	(0.0015)			(0.0000)	(0.0001)
2009–Population density			0.2633***	0.1750			0.0194	0.0206**
			(0.0981)	(0.1220)			(0.0128)	(0.0101)
Distance to coastal line			0.0077	–0.0025			0.0005***	0.0000
			(0.0071)	(0.0172)			(0.0002)	(0.0004)
Province FE		Yes		Yes		Yes		Yes
N districts	687	687	687	687	687	687	687	687
R-squared	0.239	0.584	0.368	0.621	0.056	0.242	0.203	0.365

Notes:
Other controls were the Kinh ethnic rate, nighttime light intensity, population density in 2009, urban land ratio, and cropland ratio in 1992, mean elevation in 1996, and distance to the coastline.
Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table 2 Correlations between density of present-day schools, health facilities, employees, and average distance

Density	Education							Health			
Variables	Employees (1)	Primary school (2)	Secondary school (3)	High school (4)	Vocational school (5)	College (6)	University (7)	Employees (8)	Hospital (9)	Clinic (10)	Communal health station (11)
Without province FE											
The IV – *Average distance* to the imperial test venues	–0.0000***	0.0000	–0.0000***	–0.0000***	–0.0000***	–0.0000***	–0.0000**	–0.0000***	–0.0000***	–0.0000**	–0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N districts	687	687	687	687	687	687	687	687	687	687	687
R–squared	0.366	0.542	0.613	0.178	0.146	0.150	0.369	0.565	0.136	0.047	0.609
With province FE											
The IV – *Average distance* to the imperial test venues	–0.0000	0.0000	–0.0000	0.0000	0.0000	–0.0000	–0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N districts	687	687	687	687	687	687	687	687	687	687	687
R-squared	0.472	0.728	0.684	0.312	0.338	0.354	0.492	0.719	0.266	0.289	0.755

Notes:
All estimations included the Kinh ethnic rate, nighttime light intensity, population density in 2009, urban land ratio, and cropland ratio in 1992, mean elevation in 1996, and distance to the coastline.
All outcomes were measured in density (per capita in 2009).
Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). P-weight was in all estimations.

Table 3 Effects on present-day educational achievements

Variables	Literate rate	Years of schooling			Attendance rate	Dropout rate	Never-enroll rate	Attendance rate		Taking NEEU after 12th grade
Age cohort	7+	18+	22+	25+	7-10	7-10	7-10	1991-born	18-24	1991-born
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Elite number	0.0004***	0.0402***	0.0326***	0.0320***	0.0006***	–0.0002***	–0.0004***	0.0056***	0.0017**	0.0002
	(0.0001)	(0.0067)	(0.0045)	(0.0045)	(0.0001)	(0.0000)	(0.0000)	(0.0006)	(0.0007)	(0.0006)
N districts	687	687	687	687	686	686	686	686	687	681
R-squared	0.567	0.622	0.607	0.595	0.339	0.288	0.373	0.371	0.495	0.460
IV 2nd stage	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Elite number	0.0023***	0.1399***	0.1244***	0.1247***	0.0024***	–0.0011***	–0.0014***	0.0195***	0.0069***	0.0019**
	(0.0003)	(0.0133)	(0.0121)	(0.0123)	(0.0002)	(0.0001)	(0.0001)	(0.0018)	(0.0010)	(0.0008)
N districts	687	687	687	687	686	686	686	686	687	681
F-statistics†	118.45	118.45	118.45	118.45	118.45	118.45	118.45	118.45	118.45	118.06

Notes:
† Kleibergen-Paap Wald rk F statistic for testing H0: Weak identification test.
Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). P-weight was in all estimations.
All estimations included the Kinh ethnic ratio, nighttime light intensity, population density in 2009, urban land ratio, and cropland ratio in 1992, mean elevation in 1996, and distance to the coastline as controls.

Table 4 Effects on 2009 NEEU z-scores

Classification	A				B			
Variables	Sum	Physics	Mathematics	Chemistry	Sum	Biology	Mathematics	Chemistry
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Elite number	0.0049***	0.0053***	0.0029***	0.0047***	0.0057***	0.0004	0.0071***	0.0052***
	(0.0007)	(0.0008)	(0.0006)	(0.0006)	(0.0010)	(0.0005)	(0.0011)	(0.0008)
R-squared	0.043	0.040	0.036	0.026	0.068	0.024	0.070	0.052
IV 2nd stage	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Elite number	0.0217***	0.0245***	0.0123***	0.0203***	0.0237***	–0.0015	0.0314***	0.0212***
	(0.0021)	(0.0023)	(0.0015)	(0.0020)	(0.0023)	(0.0011)	(0.0030)	(0.0022)
F–statistics†	134.23	134.23	134.23	134.23	131.72	131.72	131.72	131.72
N test takers	340,855	340,855	340,855	340,855	178,704	178,704	178,704	178,704

Classification	C				D			
Variables	Sum	Literature	History	Geography	Sum	Literature	Mathematics	Foreign language
OLS	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Elite number	0.0027***	0.0056***	0.0002	0.0014	0.0050***	0.0085***	0.0039***	–0.0005
	(0.0009)	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0014)	(0.0009)	(0.0006)
R-squared	0.038	0.092	0.019	0.020	0.079	0.096	0.060	0.084
IV 2nd stage	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
Elite number	0.0213***	0.0411***	0.0043**	0.0105***	0.0302***	0.0443***	0.0231***	0.0038**
	(0.0026)	(0.0039)	(0.0020)	(0.0021)	(0.0034)	(0.0048)	(0.0027)	(0.0015)
F-statistics†	132.19	132.19	132.19	132.19	109.34	109.34	109.34	109.34
N test takers	50,721	50,721	50,721	50,721	111,029	111,029	111,029	111,029

Notes:
† Kleibergen-Paap Wald rk F statistic for testing H0: Weak identification test.
Robust high-school clustered standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
All estimations included sex, $no\ privilege$, the Kinh ethnic ratio, nighttime light intensity, population density in 2009, urban land ratio, and cropland ratio in 1992, mean elevation in 1996, and distance to the coastline as controls.