Universidad de San Andrés

Departamento de Economía

Licenciatura en Economía

Does rice cultivation induce better math institutions?

Author: Paola Llamas (31433)
Mentor: Martín A. Rossi

Victoria (Buenos Aires), Septiembre 2023

Does rice cultivation induce better math institutions?*

Paola Llamas
Universidad de San Andrés

Abstract

This paper studies the link between historical rice cultivation and the 'math quality' of institutions in Chinese provinces. To address potential endogeneity concerns, we use Rice Suitability as an instrumental variable for rice cultivation. We find strong evidence of causal relationship between historical rice cultivation and institutions 'math quality', even after addressing potential endogeneity concerns and micronumerosity issues. Our findings suggest a novel perspective over conventional determinants of educational quality and evidence of a new potential long-term effect of rice cultivation.

Acknowledgments

I want to thank my mentor, Martín A. Rossi, for his thorough guidance and Christian Ruzzier for his support throughout this journey. Additionally, I thank researcher Thomas Talhelm for providing the data that made this study possible.

* Paola Llamas, Universidad de San Andrés, Vito Dumas 284, Victoria, Buenos Aires Province, Argentina. Email: pllamas@udesa.edu.ar. All the content is my work, and I take full responsibility for errors.

1. Introduction

When it comes to math, Asians have a built-in advantage. It is an unusual kind of advantage. For years, students from China, South Korea, and Japan have substantially outperformed their Western counterparts in mathematics (Gladwell, 2008). There is abundant data that reflects their outstanding performance in the field. In 2016, the OECD released a report showing the global rankings of student performance in mathematics. The results show that Asians have obliterated other continents: Singapore, Hong Kong, Macao (China), and Chinese Taipei were the top placers in mathematics. They analyzed the PISA 2015 exam, in which students obtain a proficiency level that ranges from 1 (lowest level) to 6 (score higher than 669 points). Across OECD countries, only 2.3% of students attain Level 6, while more than one in ten students perform at this level in Singapore and Chinese Taipei.[1]

However, how did these statistics come to be? Why are Asian students known to be prolific in the field of mathematics? Several theories have been developed in previous literature to answer this question, which are reinforced in Gladwell (2008). One of these theories refers to the linguistic structure of the Asian language. Dehaene (1997) argues that as Chinese words for numbers are shorter, these facilitate their memory and they can remember more sequences of numbers compared to western population. A second theory explored by Boe et al. (2002) refers to Asian Persistence. The researcher analyzed the TIMSS, a worldwide math and science exam for elementary and high school students. In this exam, students have to fill out an additional questionnaire of 120 questions (about their parents, education, among others), where many students leave as many as twenty questions blank. The authors discovered that if you compare the questionnaire rankings with the math rankings on the TIMSS, they are the same

[1] For further details, see OECD report.

(not related, identical). Singapore, South Korea, Taiwan, Hong Kong, and Japan are at the top of both lists. This evidences Asian persistence and further literature, such as Blinco (1993), supports the idea that they are culturally more perseverant, which is a key attribute to excel in mathematics. Another set of potential explanations for the exceptional performance is quite direct. It entails considering that Asian countries possess higher-quality educational institutions in the realm of Mathematics. Consequently, due to their access to a more profound level of education, this advantage translates into superior performance on a global scale.

This study centers on the last set of explanations and delves into the factors influencing the quality of educational institutions. Extensive research has been conducted in this area. For instance, Pietrucha (2018) explores country-specific factors that impact the Academic Ranking of World Universities, the QS World University Ranking, and the Times Higher Education World University Ranking, considered the top three global university assessments. The author concludes that the relevant variables are the economic strength of the country, its investment in research and development, stability in the long-term political arena (marked by the absence of war, occupation, coups, and significant political system changes), and institutional factors such as government effectiveness. In this context, we will focus on the determinants of the quality of educational institutions, but on the mathematical quality precisely ('math quality' hereafter).

Our main objective is to examine if rice production is a determinant of educational institutions' 'math quality'. There is a valid basis for this association. The potential positive impact of rice cultivation on the development of better mathematical institutions lies in the inherent demands and characteristics of rice farming. Historically, rice cultivation consisted of two main features: smart and hard work. Concerning the first one, as described by Bray (1994), rice agriculture is skill oriented: farmers improved their yields by becoming smarter and using every square inch of the rice paddy. In the second place, historically, people who grew rice have worked harder than almost any other kind of farmer. In *The Discovery of France*, Graham Robb argues that in the Pyrenees and the Alps, entire villages would hibernate from the time of the first snow in November until March, while if you were a peasant farmer in Southern China, you did not sleep through the winter, instead busied yourself with side tasks (Gladwell, 2008).

In sum, rice farmers had to optimize yields and effectively manage limited land resources, requiring strategic planning and problem-solving skills. This environment of diligent and intelligent agricultural practices could have fostered a mindset that appreciates precision, analytical thinking, and efficient resource management—attributes that align closely with the skills needed for success in mathematics. Thus, the challenging nature of rice cultivation may have indirectly catalyzed the development of a solid mathematical foundation within communities historically involved in this agricultural practice. This, in turn, could have potentially positively influenced the emergence of superior mathematical institutions within these regions.

Therefore, our central hypothesis is that historical rice cultivation has had a positive impact on the 'math quality' of institutions. Our contribution has two essential components. First, while previously identified determinants of high-quality institutions generally apply to various fields, there remains a notable gap in research concerning the determinants of 'math quality' specifically. Furthermore, the link between rice production and 'math quality' has not been previously considered or examined. Therefore, we suggest a novel perspective over conventional determinants of educational quality. This holds relevance as countries currently transitioning their agricultural practices to rice cultivation could be exposed to a potential long-term advantage they are unaware of. The shift towards rice farming could promote sustained academic prosperity in the long run.

Our empirical approach initially uses Ordinary Least Squares (OLS) to estimate the connection between *Average Ranking* and *Percent Paddy*, that serve as our measure of 'math quality' and rice cultivation respectively. To address potential endogeneity, we use *Rice Suitability* as an instrumental variable for *Percent Paddy*. The findings consistently reveal a significant negative effect, indicating that provinces with greater paddy rice cultivation tend to have higher-ranking mathematics institutions. These results are robust when examining an alternative representation of rice cultivation and employing Bootstrap standard errors.

The structure of this paper unfolds as follows. Section II describes the data. Section III presents the empirical strategy and the results. Section IV summarizes the main results, mentions limitations, and inspires further explorations in this domain.

2. Data

We use province-level data from China. In this section, we present the measure of rice cultivation, the measure of 'math quality,' and all additional variables used in the empirical strategy.

The measure of rice cultivation

There are two considerations when considering an adequate measure for rice cultivation. First, we want to measure historical rice rather than contemporary data. This is meant to record the traditional farming practices of different regions, rather than data influenced by recent advancements in irrigation and machinery. Also, our research focuses on historical patterns since it is reasonable to think that the intervention of rice cultivation does not have an immediate impact on the outcome of interest. This is pivotal because it refers to the maturity of the intervention. China presents a unique case to "test" the theory as it has a long history of rice cultivation, giving enough time for differences in the quality of institutions to play out. Second, rice cultivation has different variants. We use data on paddy rice type specifically rather than aggregating all types. For instance, dryland rice grows naturally, without the need for constant attention and hard work present in paddy rice. Therefore, given the mechanism we have in mind, we expect to observe more substantial effects when focusing exclusively on the paddy rice variant.

For this objective, we follow Talhelm (2020) and use 1996 statistics from the China Statistical Yearbook. This large-scale yearbook contains statistical information comprehensively reflecting economic and social development. A section indicates for each province the total area of cultivated land and the area that corresponds to paddy rice fields. Thus, our primary explanatory variable is *Percent Paddy*, the percentage of cultivated land devoted to paddy rice. As stated in Talhelm's (2020) research, these data strongly correlate with the available 1918 data for a subset of 22 provinces. Therefore, 1996 data accurately reflects historical patterns of rice farming. With this data, we also construct a dummy variable *Majority Rice* that takes value one if *Percent Paddy* > 50 %, as done in Talhelm (2020), which will allow us to consider different econometric specifications. For simplicity, in this study, we describe non-rice farming regions as 'wheat-farming' regions. This is a simplification because non-rice regions also traditionally grew similar dryland crops like millet and barley.

Measuring the 'math quality' of institutions

Scimago Institutions Rankings is a website that ranks academic and research institutions based on various criteria. These rankings are often used to assess universities' and research organizations' research output and impact worldwide. Scimago uses data from the Scopus database to generate these rankings, which include metrics such as research output, citations, international collaboration, and more. We use this website to put together a ranking of math institutions for each province of China. To do this, we started by narrowing down and ranking universities in China based on the field of Mathematics.[2] This gave us a list of 465 ranked institutions from all over the country. Then, we created a database that links each institution to its respective province. Finally, we calculated the average ranking for each province, ensuring that each one had at least one institution represented in the rankings. We label this variable as *Average Ranking*, constituting the outcome of interest throughout the work.

Additional data

As it will be described in the following section, *Percent Paddy* and *Majority Rice* can be endogenous. Therefore, to gain credibility in the internal validity of the results, we aim to exploit exogenous variations that determine regional differences in rice farming. For this, we have data on *Rice Suitability*. This is an index created by the United Nations Food Agriculture that estimates the environmental suitability of wetland rice for each province based on temperature, ground slope, soil, among other variables, from 1961 to 1990. This variable will be used as an instrument for rice cultivation to estimate the parameter of interest via 2SLS.

In addition, we have data on province GDP per capita for the years 1996, 2000, 2008, 2012, and 2014, measured in yuans. We will primarily use the most recent data from 2014, which we name as *GDPPC*. Additionally, we possess data on 150 other variables that vary among provinces. However, since we expect most of these variables to have little to no correlation with either *Percent Paddy* or *Average Ranking*, we do not include them in our econometric analysis. Table 1 presents summary statistics of the data.

3. Empirical Strategy and Results

[2] See Scimago Ranking.

We are interested in estimating the causal effect of rice cultivation on the 'math quality' of institutions in China. Formally, we estimate:

$$Average\ Ranking_i = \beta_0 + \beta_1 Percent\ Paddy_i + \beta_2 X_i + \mu_i \quad (1)$$

where i index provinces. The dependent variable is $Average\ Ranking_i$, and the explanatory variable is $Percent\ Paddy_i$. The parameter of interest is β_1, and X_i is the set of province-level control variables. In the majority of specifications, X_i includes only province GDP per capita. Last, μ_i is an error term.

Given that there are thirty-one provinces, for all estimates we report bootstrap standard errors, additional to robust standard errors. The results are shown in Table 2. In Column (1), we report OLS estimates on the relationship between *Average Ranking* and *Percent Paddy* without including controls. In Column (2), we report OLS estimates on the relationship between *Average Ranking* and *Percent Paddy*, including province GDP per capita as control.

In both models, *Percent Paddy* has a negative and significant coefficient. The estimated coefficients indicate that when the percentage of rice cultivation dedicated to paddy rice rises by one percentage point, the Average Ranking decreases by 32 points and it is significant at the 10% level. Thus, provinces with a greater proportion of land devoted to paddy rice production are associated with better overall rankings (= less ranking number) for their educational institutions within that province in mathematics. Last, as expected, the province's GDP per capita has a negative and significant coefficient. When GDP per capita rises by one Yuan, the ranking decreases by -0.0004 points on average, and the effect is significant at the 5% level. Thus, more prosperous provinces are associated with a better ranking in their math institutions, on average, and *ceteris paribus*.

Percent Paddy, however, may be endogenous in a model for *Average Ranking*. For example, provinces where the government invests more in promoting and subsidizing rice cultivation may also allocate resources to improve education. To address potential endogeneity concerns, we instrument *Percent Paddy* with *Rice Suitability*.

The first stage is shown in Table 2. Column (3) show the first stage estimates without additional controls. Column (4) reports the first stage adding GDP as control. In both cases, *Rice Suitability* is significant in explaining *Percent Paddy* at the 1% level. The F-statistic is equal to 67 when including controls, well above the threshold for weak instruments. This provides strong evidence for the relevance of the instrument.

The identification assumption is that *Rice Suitability* is not directly related to *Average Ranking* after conditioning on the set of controls. As detailed in the Data section, the index of *Rice Suitability* is based on variables such as ground slope and humidity. It is reasonable to think the exogeneity assumption holds since the ground slope, for instance, does not appear to directly impact the 'math quality' of institutions of a particular province. Thus, differences in ground slope and humidity (among other factors weighted in the index) constitute exogenous variations that determine regional differences in rice farming.

We further estimate Equation (1) using 2SLS, and the results are shown in Table 2. In Column (5), we report 2SLS estimates on the relationship between *Average Ranking* and *Percent Paddy*. In Column (6), we report 2SLS estimates on the relationship between *Average Ranking* and *Percent Padde,* including province GDP per capita as a control variable. As in this procedure we only use the variability of rice cultivation contained in the instrument, it is likely that this represents exogenous variability of rice cultivation.

The estimated coefficients indicate that a rise of 1 percentage point of land dedicated to paddy rice causes the Average Ranking to decrease by 37 points and is significant at the 10% level. Last, the province's GDP per capita has a negative and significant coefficient, and the conclusions are identical to the specification without instrumenting. All results remain robust when considering Bootstrap Standard Errors.

For robustness, we follow the same procedure considering *Majority Rice* as the explanatory variable. The results are shown in Table 3. In Column (1), we report OLS estimates on the relationship between *Average Ranking* and *Majority Rice* without including controls. In Column (2), we report OLS estimates on the relationship between *Average Ranking* and *Majority Rice*, including province GDP per capita as control. All coefficients maintain sign and significance. To address endogeneity concerns, we instrument *Majority Rice* with *Rice Suitability*. The first stage is shown in Column (3) and (4). In both cases, *Rice Suitability* is significant in explaining *Majority Rice* at the 1% level. Finally, 2SLS estimates are shown in Column (5) and (6). The results show that rice provinces are associated with an average ranking of 26 points lower than wheat provinces, all other factors being equal and significant at the 5% level. All results remain robust when considering Bootstrap Standard Errors.

4. Conclusion

Our study, motivated by the exceptional mathematical performance of Asian countries, has provided insights into the role of historical rice production as a determinant of educational institution quality. We specifically examined the causal impact of historical rice cultivation on the 'math quality' of educational institutions in Chinese provinces.

Our findings provide strong evidence of a significant and causal relationship between historical rice cultivation and institutions 'math quality', even after addressing potential endogeneity concerns and micronumerosity issues. This relationship is elucidated through our proposed mechanism: the demanding nature of rice farming, necessitating precision, analytical thinking, and efficient resource management, played a pivotal role in fostering robust mathematical foundations within communities historically involved in this agricultural practice.

In unveiling this link, this study makes two distinct yet interconnected contributions to educational research. Firstly, it offers an innovative perspective by introducing rice production as a previously unexamined determinant of educational excellence. Certainly, this challenge traditional notions of what shapes the quality of education. Secondly, our research narrows its focus to 'math quality,' a dimension that has received limited attention in prior literature. Together, these aspects fill a crucial gap in the existing body of knowledge.

Our research suggests that regions transitioning to rice cultivation unlock a potential long-term advantage in academic prosperity. While we have primarily focused on Chinese provinces, future research should explore the generalizability of these findings to other regions and formally investigate the mechanisms through which rice cultivation influences ‘math quality’. Yet, our core message transcends geographical boundaries. We are emphasizing the significance of unique historical factors when analyzing regional disparities in educational performance. In the case of China, the linked factors are rice cultivation and ‘math quality’, but in every region, there is the potential for specific historical practices to shape development. These activities may encompass playing traditional card games, engaging in artisanal craftsmanship, or preserving cultural rituals. This broader perspective invites further exploration into the legacies of different regions and challenges the conventional focus on well-known determinants of educational development. Thus, our work opens the door to a deeper understanding of the interplay between culture, history, and education.

5. References

Blinco, Priscilla Mary Anne. "Persistence and education: A formula for Japan's economic success." *Comparative Education* 29.2 (1993): 171-183.

Boe, E. E., May, H., Barkanic, G., & Boruch, R. (2002). Predictors of National Differences in Mathematics and Science Achievement: Data From TIMSS for Eighth Grade Students.

Bray, F. (1994). *The rice economies: technology and development in Asian societies*. Univ of California Press.

China explained by rice versus wheat agriculture. *Science*, *344*(6184), 603-608.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown.

Gurria, A. (2016). PISA 2015 results in focus. *PISA in Focus*, (67), 1.

Pietrucha, J. (2018). Country-specific determinants of world university rankings. *Scientometrics*, *114*(3), 1129-1139.

Robb, G. (2016). *The Discovery of France: Picador Classic* (Vol. 41). Pan Macmilla

State Statistical Bureau of the People's Republic of China, China Statistical Yearbook (1996, 2005).

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. Science, 344(6184), 603-608.

Talhelm, T., & English, A. S. (2020). Historically rice-farming societies have tighter social norms in China and worldwide. Proceedings of the National Academy of Sciences, 117(33), 19816-19824.

Zhu, C., Talhelm, T., Li, Y., Chen, G., Zhu, J., & Wang, J. (2021). Relationship between rice farming and polygenic scores potentially linked to agriculture in China. *Royal Society open science*, *8*(8), 210382.

Table 1. Summary statistics

	Observations	Mean	Standard Deviation	Minimum	Maximum
Percent Paddy	31	.33	.31	0	.88
Majority Rice	31	.39	.50	0	1
Average Ranking	30	171.63	31.53	122.03	242.67
Rice Suitability	31	24.81	23.95	0	56.2
GDPPC	31	50,742.77	22,080.56	26,433.00	10,5231.00

Table 2. Results

	(1) Average Ranking	(2) Average Ranking	(3) Percent Paddy	(4) Percent Paddy	(5) Average Ranking	(6) Average Ranking
Percent Paddy	-36.58	-32.91			-34.76	-37.62
	(17.53)**	(16.69)*			(20.97)*	(19.37)*
	[16.94]**	[16.60]**			[21.77]	[21.21]*
GDPPC		-0.000482		2.44e-06		-0.000475
		(0.000207)**		(9.83e-07)**		(0.000196)**
		[0.000223]**		[1.08e-06]**		[0.000225]**
Rice Suitability			0.0114	0.0115		
			(0.00121)***	(0.00114)***		
			[0.00118]***	[0.00113]***		
Observations	30	30	31	31	30	30
R-squared	0.133	0.245	0.752	0.781	0.132	0.243
F-statistic	-	-	88.84	67	-	-

Note: Robust standard errors are in parentheses and Bootstrap standard errors (1000 repetitions) are in brackets. *Significant at the 10% level. **Significant at the 5% level. **Significant at the 1% level.

Table 3. Results: Majority Rice

	(1) Average Ranking	(2) Average Ranking	(3) Majority Rice	(4) Majority Rice	(5) Average Ranking	(6) Average Ranking
Majority Rice	-28.39	-25.90			-24.52	-26.47
	(9.676)***	(9.007)***			(14.32)*	(13.08)**
	[9.351]***	[8.917]***			[15.10]	[14.49]*
GDPPC		-0.000462		3.88e-06		-0.000460
		(0.000200)**		(2.10e-06)*		(0.000187)**
		[0.000215]**		[2.29e-06]*		[0.000222]**
Rice Suitability			0.0160	0.0162		
			(0.00209)***	(0.00200)***		
			[0.00212]***	[0.00209] ***		
Observations	30	30	31	31	30	30
R-squared	0.201	0.304	0.602	0.632	0.198	0.304
F-statistic	-	-	58.91	80.98	-	-

Note: Robust standard errors are in parentheses and Bootstrap standard errors (1000 repetitions) are in brackets. *Significant at the 10% level. **Significant at the 5% level. **Significant at the 1% level.