Journal of Development Economics 176 (2025) 103506

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

The deeper roots of human capital formation and economic development in Southeast Asia, 1900–2000

Alexandra M. de Pleijt [a,*], Ewout Frankema [a,b,*,1]

[a] Wageningen University, The Netherlands
[b] CEPR, United Kingdom

ARTICLE INFO

JEL classification:
J24
O10
O40
O53
N00

Keywords:
Human capital
Southeast Asia
Economic development
Economic history
Modern economic growth

ABSTRACT

Since 1970, Southeast Asia's per capita GDP grew seven-fold and headcount poverty rates declined from ca. 70 to 5%. This paper explores the 20th century schooling revolution as one of the deeper roots of this major leap in human prosperity. Using micro-data on the educational attainment and migration status of ca. 123 million individuals, subdivided across 277 provinces in eight Southeast Asian countries, we establish a strong and significant relationship between early educational attainment and sub-national economic development at the start of the 21st century. Using a wide range of historical and geographic controls, we find that higher education shares are more strongly associated with regional development outcomes than mass education. We also find a strong and robust contribution of inter-regional and international migration to human capital accumulation and long-term development.

1. Introduction

Southeast Asia experienced impressive economic growth and poverty reduction over the past half century. Since 1970, the region's aggregate per capita GDP rose seven-fold. The share of people living in extreme poverty declined from ca. 70% to 5% and the share of agriculturally employed in the region's labour force decreased from over 80% in 1950 to ca. 30% in 2022 (ILO, 1997; GGDC, 2014; World Bank, 2024). The proximate causes of this major leap in human prosperity are well understood: rapid labour productivity growth in all main sectors of the economy underpinned by high rates of domestic saving, capital formation and concomitant technological and organizational innovations in a broader context of improving macroeconomic stability. These innovations, in turn, were inconceivable without substantial public and private investments in human capital formation (World Bank, 1993; Nayyar, 2019). In this regard, Southeast Asia's stellar growth performance sits well with the emphasis placed by the New and Unified Growth Theory on human capital as the ultimate driver of the transition to modern economic growth (cf. Nelson and Phelps (1966), Mankiw et al. (1992), Goldin and Katz (1998), Mokyr (2002) and Galor (2011)).

But how deep are the roots of Southeast Asia's economic 'take-off'? According to one strand of literature not very deep. This literature stresses the role of growth-promoting, pro-poor policies adopted by various national governments during the post-colonial era, and which helped kick-start growth in one country after another, albeit with different velocities (Bevan et al., 1999; Berendsen et al., 2013). These policies were especially aimed at rural development and helped spread the productivity advances of the Green Revolution, thus laying the foundations for subsequent industrialization and structural change (Lipton, 1977; Henley, 2012, 2015). This literature emphasizes how poor and underdeveloped Southeast Asian economies were prior to 1970, how their sector structures barely differed from Sub-Saharan African

* Corresponding authors at: Wageningen University, The Netherlands.
E-mail addresses: sandra.depleijt@wur.nl (A.M. de Pleijt), ewout.frankema@wur.nl (E. Frankema).

[1] For comments on earlier drafts we thank Steve Broadberry, David de la Croix, Leigh Gardner, Michiel de Haas, Gregg Huff, Micheal Montesano, as well as the editor Andrew Foster and two anonymous reviewers. We also thank seminar participants at Humboldt University, Yale University, Utrecht University, Wageningen University, the World Cliometrics Conference 2023 (Dublin), the European Historical Economics Society Conference 2023 (Vienna), the Association for Asian Studies Conference 2023 (Boston) and the CEPR 40th Anniversary Symposium 2023 (Paris). For excellent research assistance we thank Jakko-Jan van Ek, Annemieke Verhoek and Lucas Koren. We gratefully acknowledge financial support from The Netherlands Organisation for Scientific Research for the project "South-South Divergence: Comparative Histories of Regional Integration in Southeast Asia and Sub-Saharan Africa since 1850" (NWO VICI Grant Number VI.C.201.062).

https://doi.org/10.1016/j.jdeveco.2025.103506
Received 20 August 2024; Received in revised form 24 March 2025; Accepted 24 March 2025
Available online 5 April 2025

economies, and how successive crises, violence, political disorder and isolationism caused economic stagnation between the 1930s and 1970s. Indeed, few if any contemporary observers foresaw the spectacular economic transition that waited just around the corner (cf. Geertz (1963) and Myrdal (1968)).

In contrast, economic historians have argued that Southeast Asia's 'miracle' growth can be traced back into the colonial era, well before the period of instability and disintegration. Despite the many missed development opportunities of particular colonies (see Booth (1998) for Indonesia) or post-independence states (see Brown (2013) for Burma), this literature has so far identified four deeper roots: (1) indigenous and colonial investments in (extractive) commercial agriculture (Booth, 1998; Dell and Olken, 2020); (2) progressive integration of labour and food (rice) markets since the late 19th century (Huff and Caggiano, 2007; Marks, 2010); (3) expansion of proto-industrial activities (Huff and Angeles, 2011; van Nederveen Meerkerk, 2017) and; (4) a general rise in real wages and living standards which have been interpreted as Southeast Asia's 'first' economic miracle (Bassino and der Eng, 2021).

Exploring the long-run development effects of 20th century human capital formation, this paper adds a new dimension to this discussion, one that can bridge the two perspectives on the roots of modern economic growth in Southeast Asia. Booming export sectors during the colonial era required brokers, clerks, bankers and lawyers to co-ordinate extended value chains and increased contractual complexity. Engineers, technicians and skilled construction workers were needed in increasing numbers to build commercial infrastructures. As Henderson et al. (2018) have shown, geographic predispositions to trade explain a much larger part of the variation in current nightlight density levels of developing economies, then they do in early industrializing countries where conditions for agriculture were key. The long-run effects of trade, in turn, may have been transmitted via various channels including persistent agglomeration effects, induced innovations of market institutions, investments in physical infrastructure and proto-industry, and indeed, through increased demands for skilled labour.

At the same time, human capital was also essential for the design and implementation of pro-rural policies that underpinned the diffusion of productivity-enhancing agricultural technologies in the post-colonial era, as well as rural welfare development more broadly. To be effective, such policies required the kind of administration, coordination and communication skills that are conducive to well-functioning state bureaucracies, and which are a central object of study in the 'developmental states' literature (Johnson, 1982; Kohli, 2004). In other words, it would not only be interesting to trace the roots of human capital formation back into the colonial era to understand its role in this early phase of expanding export trade, it would also be useful to study the trends during the period in which large parts of Southeast Asia experienced a prolonged setback.

This paper presents new evidence on the *formation* and *accumulation* of human capital across 277 Southeast Asian provinces from the 1910s through to the 1960s. Focusing on the formative stages of its modern education system, we add the understudied case of Southeast Asia as a 'late' developing region, to a larger comparative literature on the education-growth nexus. While the role of formal education in the Industrial Revolution of Western Europe and the United States remains heavily debated, studies of Asian industrialization have been in broad agreement that a formally schooled workforce was crucial for catch-up growth (Amsden, 1989; Birdsall et al., 1997; Krueger and Lindahl, 2001; Stubbs, 2018). Yet, so far, these studies have largely relied on post-1950 national-level schooling data (Benhabib and Spiegel, 2005; Hanushek and Woessmann, 2011; Pritchett, 2013; Barro and Lee, 2015). In their landmark study on human capital in the long-run, Lee and Lee (2016) present time-series going back to the 1820s, but their analysis hinges on national-level data and takes a global comparative perspective. We exploit the fine-grained micro-data on educational attainment and migrations from population censuses made available by IPUMS international (Ruggles et al., 2024), to trace the chronology of Southeast Asia's schooling revolution back into the early 20th century at the *sub-national level*. Our study includes data for Cambodia, Indonesia, Laos, Malaysia, Papua New Guinea, the Philippines, Thailand and Vietnam.[2]

Human capital is a broad concept, comprising health, age, cognitive, motoric and social abilities, creativity and knowledge, and is therefore notoriously hard to measure. Following Lee and Lee (2016) we focus on formal education as the central channel of human capital accumulation in modern economies. We highlight four main findings. First, regression analyses suggest a strong and significant positive association between early 20th century attainment levels and early 21st century development outcomes. Following Henderson et al. (2018), we use nightlight density levels as a proxy for sub-national economic development. This association is robust to the inclusion of a large set of confounding variables, including controls for early urbanization and trade geographies, as well as alternative specifications of our models (a.o. adjustments for spatial autocorrelation). This finding suggests that the schooling revolution initiated during the colonial era did have a lasting long-term impact on spatial inequalities in economic development.

Our second contribution pertains to the distinction between mass education and higher education. With regards to the First Industrial Revolution in Britain, it has been argued that mass education was rather unimportant as it were small groups of highly skilled and mobile artisans who made the difference (Mokyr, 2005; Mokyr and Voth, 2009; Kelly et al., 2023). In the 19th-century US, the growth of manufacturing has led to de-skilling of the labour force, creating new jobs for the lower educated as well as higher educated, at the expense of the middle segment (Katz and Margo, 2014). Others have argued that primary schooling was important in economies catching up with Britain, such as 19th century Prussia, but that its role differed per industry. For the expansion of textile industries in Germany, which made intensive use of child labour, mass education was unimportant, while for industries such as metals, food, rubber and paper elementary and middle education was of major importance, while upper-secondary and higher education had no impact (Becker et al., 2011). For France, it has been argued that the shift in technological complexity during the 19th century required a shift from basic to intermediate levels of education (Diebolt et al., 2021).

Such questions are yet to be addressed for 'late' industrializing economies, where the gap with the global technology frontier was (much) larger than in 19th century Prussia or France. Did catch-up growth in Southeast Asia benefit from a broad bottom of the education pyramid, or were concentrated pockets of highly educated workers crucial? The results from our analysis suggest that in the context of 'late' industrialization, higher education was particularly important. Even though higher education cannot exist without a minimum level of basic education, and attainment data fail to capture aspects of educational quality, the association we find is so strong and the set of controls we use so diverse, that this finding may be considered as an anchor point in the broader debate about the long-term legacies of colonial rule and for future research exploring these associations in greater detail.[3]

Third, tracing the roots of Southeast Asia's post-colonial growth acceleration back to the emergence of relatively small, but critical minorities of highly educated workers in the colonial era is not straightforward in light of the severe economic and political crises that plagued the region during the middle decades of the 20th century. Starting with the Great Depression of the 1930s and followed by the Japanese occupation during World War II, and subsequent independence wars and civil wars from the late 1940s to 1970s, all Southeast Asian

[2] Burma is the only major country of the Southeast Asian mainland which we had to exclude as it lacks the required census data.

[3] All of this is not to say that basic education was unimportant for trans-generational processes of social development or gender emancipation, e.g. Hoang and Nguyen (2023).

economies experienced a setback in trade and income levels at some point in time. We show that investments in mass education as well as higher education continued unabated during these 'lost decades' and that the relationship between (increasing) educational attainment rates and nightlight density holds for all decadal birth-cohorts from the 1910s to 1960s.[4] This continued expansion is consistent with evidence of falling skill-premiums in the region during the middle decades of the 20th century (Frankema and Waijenburg, 2023).

Finally, the relationship we find is not only robust to the inclusion of migration-adjusted educational attainment variables, but is also strengthened when taking inter-regional and international migration into account. Combining information about place of birth (origin) with place of residence (destination) we calculate the share of migrants for each province, and compare their average schooling levels with those of the non-migrant (sedentary) share of the population. This way, we estimate the contribution of migrants to regional concentrations of human capital as well as their contribution to the aggregate (national) economy. We find that migrants are, on average, adding 2.5 to 3.5 percentage points to the nation-wide shares of higher educated on top of the shares attained by the 'sedentary' population. We also find that the reallocation of cognitive, technical and/or social skills to the economically most dynamic regions strengthens the statistical association with early 21st century nightlight density.

A note on causality is in order. Our study best qualifies as a comparative analysis of a historical (i.e. long-run) process. It aims to provide a better understanding of the process uncovering long-term persistent correlations, rather than employing strategies to isolate the causal effect of the early 20th century schooling revolution on later economic development. As we show in Section 2, the historical roots of formal schooling in Southeast Asia were extremely diverse and hard to capture in a catch-all proxy. Moreover, since human capital is mobile and skilled labour migration was common across the region, any exogenous source of spatial variation runs into the problem that locations of human capital *formation* are not necessarily locations of human capital *accumulation*. Fig. 1 visualizes the feedback loops inherent to a path-dependent process of human capital accumulation that is inextricably intertwined with long-run economic development: clustering of human capital is a response to, as well as a pre-condition for, labour productivity growth and structural change. Occupational skill-premiums, as the literature on rural–urban migration has extensively documented, create migration-inducing opportunity gaps (Bairoch and Goertz, 1986; Glaeser, 1999; Young, 2013). Migratory selectivity into economically dynamic regions complements the local *formation* of human capital. We offer evidence that such reinforcing dynamics were part of Southeast Asia's development trajectory, by using a large set of controls, a range of robustness checks and by relying on considerable temporal distance. After all, causality operates at different scales: one is the effect of a single driver (i.e. education) on long-term growth. Another scale involves the dynamic paths which can be compared across spatial-administrative entities, but in which endogeneity is taken for granted. The latter is what we aim for.

The remainder of this paper is structured as follows. Section 2 discusses the historical context of the schooling revolution in Southeast Asia and highlights some of the main reasons for the variation that we document between and within countries based on secondary literature. Section 3 introduces the census data, our dependent variable and our set of controls. Section 4 specifies the regression model and presents the main results of our empirical analysis of the schooling-development nexus. It also focuses on the role of inter-regional migration. Section 5 concludes, and our (online) appendix provides robustness checks.

[4] The exception to this rule is the stagnation in expansion of higher education in Vietnam during the protracted period of warfare between 1946 and 1975, but which later results in high rates of adult education (more on this below).

2. The schooling revolution in Southeast Asia

Formal education, in which transfers of skills and knowledge are organized in a school system with a pre-defined annual curriculum, and rules of engagement regarding attendance, examination and promotion, had existed in Southeast Asia long before the onset of Western colonial rule. Such schools were either organized by religious institutes (e.g. Islamic, Confucianist or Buddhist) to promote the spread and sustenance of specific belief systems and moral traditions, or by states and courts to nurture an educated elite class of civil servants (Wyndham, 1933). Institutionalization of knowledge transfers was also promoted by guild-like societies who required the transmission of artisanal or commercial skills in order to sustain their trades, but these usually took the form of apprenticeships where working skills were acquired at the shopfloor. Formal mass education, in which a more generic and partly secularized approach to human learning was promoted, including the training of basic skills such as reading, writing and algebra, thus emerged out of multiple grass-roots initiatives: the establishment of (communitarian) village schools, the spread of Christian missionary and Islamic schools, the expanded presence of Buddhist monasteries as well as colonial investments in public education. Higher forms of schooling, including technical, vocational and scientific education, developed at a slower pace, was obviously more concentrated and tended to have a more direct involvement of the state.

The schooling revolution in Southeast Asia started in the late 19th century and gained momentum in the first half of the 20th century. Colonial governments influenced both the speed and the quality of mass education through the public funds they allocated to different types of schools, and through the rules they set for local communities as well as foreign missionaries to establish new schools. In the course of time, such rules were also drafted for examination standards, teacher qualifications, compulsory school attendance and so on. However, the pace of educational expansion varied largely across the region. This section provides background to understand some of the key patterns, thereby using some of the data that we introduce in more detail in Section 3.

2.1. Variation between countries

Figs. 2 and 3 chart the variation in the adult population (aged 20–90) that had enjoyed at least some formal schooling at the time the census was taken, for the birth-cohorts (BC henceforth) of the 1920s (1925 henceforth) and 1960s (1965 henceforth) respectively. The maps show that among the 1925BC in Papua New Guinea just 10% had gone to school, while in the neighbouring Philippine islands over 80% had. They also show that attainment levels in Thailand and Vietnam were far ahead of Laos and Cambodia. Whereas the 1945BC in the Philippines, Thailand and Vietnam approached full primary school attainment, in Indonesia and Malaysia this took several decades longer.

Similar variation can be observed in higher levels of education. Figs. 4 and 5 show the regional shares of the adult population born in respectively the 1920s and 1960s that had completed secondary education. The maps reveal that the Philippines were clearly leading in the provision of higher education. Malaysia, Thailand and Indonesia caught-up in the early post-colonial era. In Vietnam the expansion of higher education stagnated during the decades of the Indochinese and Vietnam war (1946–1975), but remarkable progress was made in the spread of mass education, including adult-education (Woodside, 1983).

These between-country differences in attainment levels were, to a considerable extent, the result of different colonial policies. Colonial states required educated workers in particular for expanding government administrations and export industries. State administration requires bookkeeping, tax assessments and collection, post and telegraph services, and more in general the processing of quantitative and qualitative information. State-funded education in the 19th century catered primarily to training a class of civil servants. Local elites,

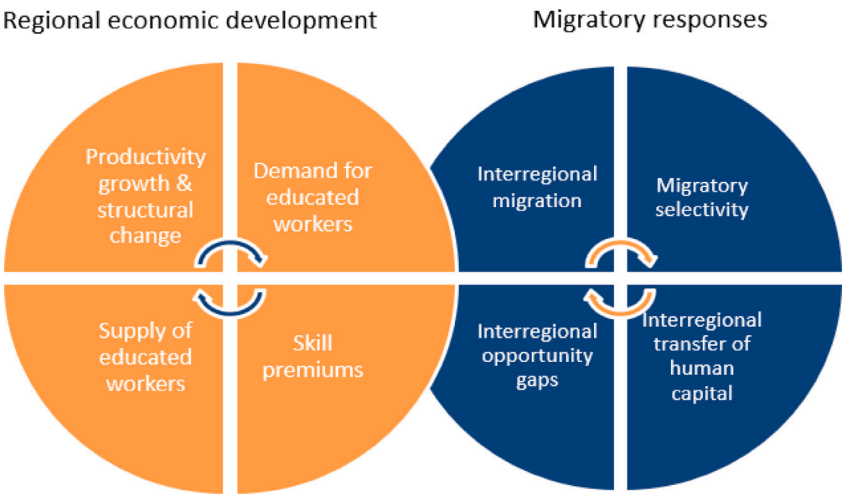

Fig. 1. Feedback loops in regional economic development and interregional migration.
Source: Authors' own.

in turn, demanded schooling in order to secure their position in the public administration (Altbach and Umakoshi, 2004; Fry, 2018). The provision of technical and commercial education expanded later, and largely in response to the rapid growth of the export economies towards the end of the 19th century when globalization spurred by railways, steamships, telegraphs and free trade agreements kicked-in (O'Rourke and Williamson, 1999). Export sectors not only required workers with commercial and administrative skills, but also enhanced demand for people with specialized technical, construction and engineering skills to build and maintain factories, railways, warehouses, machinery, tools, ships and so on.

Increasing demand for skilled workers was accommodated in four ways: (1) Immigration from the metropole. This was a costly solution in terms of transportation and salary costs (premiums), but such labour migration occurred throughout the colonial era; (2) Immigration from the region. British Malaya is a key example. Furnivall (1941, p. 28) discusses how the creation of a federal administration for the Malay states relied on Englishmen for positions of 'trust and responsibility', while clerks were recruited from Ceylon or the Straits, policemen from India or local Malay countries and railway, postal and telegraph officers from India and Ceylon (See also Kaur (2004) ; and for Indonesia see Hup and Zwart (2025)); (3) Send children of local elites abroad for education; and (4) Invest in the expansion of local education. This solution required a long time-horizon and higher upfront investments, but would eventually result in a broader and more reliable supply of skilled labour.

Colonial governments held different and changing views on these policy options. On the one hand the local provision of education underscored the legitimacy of colonial rule and its moral commitment to 'civilization' and 'development'. The US-controlled colonial government in the Philippines saw schooling as a pre-requisite for the transfer of sovereignty, which was deemed the ultimate objective of US 'trusteeship' over the archipelago (Furnivall, 1943). However, Dutch and French conceptions of the desirability of educational expansion in respectively Indonesia and Indochina were more conservative, and the funds that were made available to support it were correspondingly smaller (Booth, 2007; Frankema, 2013). Alternatively, until the depression of the 1930s, the Dutch did commit to an 'open door' policy to skilled Chinese labour immigrants who were easier to control (Hup and Zwart, 2025). Colonial officials feared that educating local populations could destabilize the social compound underpinning colonial societies. There were also concerns among French colonial officials that the spread of 'traditional' schools in Vietnam, outside the purview of the colonial state, could become a breading ground for anti-colonial activism (Kelly, 1977). That such fears were not unjustified has been testified by the fact that leaders of independence movements often came from the higher educated strata of the indigenous population.

It would be undue though to depict Southeast Asia's schooling revolution exclusively as a function of different colonial policies. The Thai administration, which had managed to keep European imperialists at arm's length, followed a comparable trajectory of educational modernization. While the country opened up (under British pressure) to trade with Western societies, and allowed the influx of Christian missionaries, it adopted Western teaching methods which gradually replaced those of by the Buddhist temple schools (Ingram, 1971; Fry, 2018). Moreover, indigenous demand for schooling, and local initiatives to supply it, were equally if not more important for the spread of formal education than top-down supply. While indigenous elites demanded schooling opportunities to acquire positions in the (colonial) state apparatus, both villagers and city-dwellers developed a keen interest in formal education in view of growing opportunities of social mobility in commerce, crafts and lower-tier clerical jobs. Local demands for education were also part of a long-term process of cultural change, anti-colonial resistance and encroaching Socialist ideology. As the skills to read, write and calculate were associated with social prestige, observing more people in local communities acquire such skills rose incentives to emulate. Many schools were thus established without any (financial) aid from, or approval by, a central colonial government. Moreover, primary education was mostly provided in the local vernacular, or the dominant language of immigrant communities, which differed from the situation in many African countries where the language of the colonizer became the main language of instruction (Laitin and Ramachandran, 2022).

2.2. Variation within countries

Despite its eminent importance, economic research on the unequal diffusion of mass education within Southeast Asian societies has remained surprisingly thin. In Thailand and the Philippines, for instance, the share of the 1925BC that went to school ranged from ca. 20% to over 90%. In Indonesia large differences occurred as well, albeit around a lower mean. In Papua New Guinea, where the schooling revolution had barely taken-off, these spatial disparities were more compressed, with one notable outlier, the tiny capital city Port Moresby. The patterns that emerge from the micro-data shown in Figs. 2 to 7 raise various pertinent questions: why did some regions advance much faster than others? How important were these early regional pockets of educational growth for later processes of structural change and economic development?

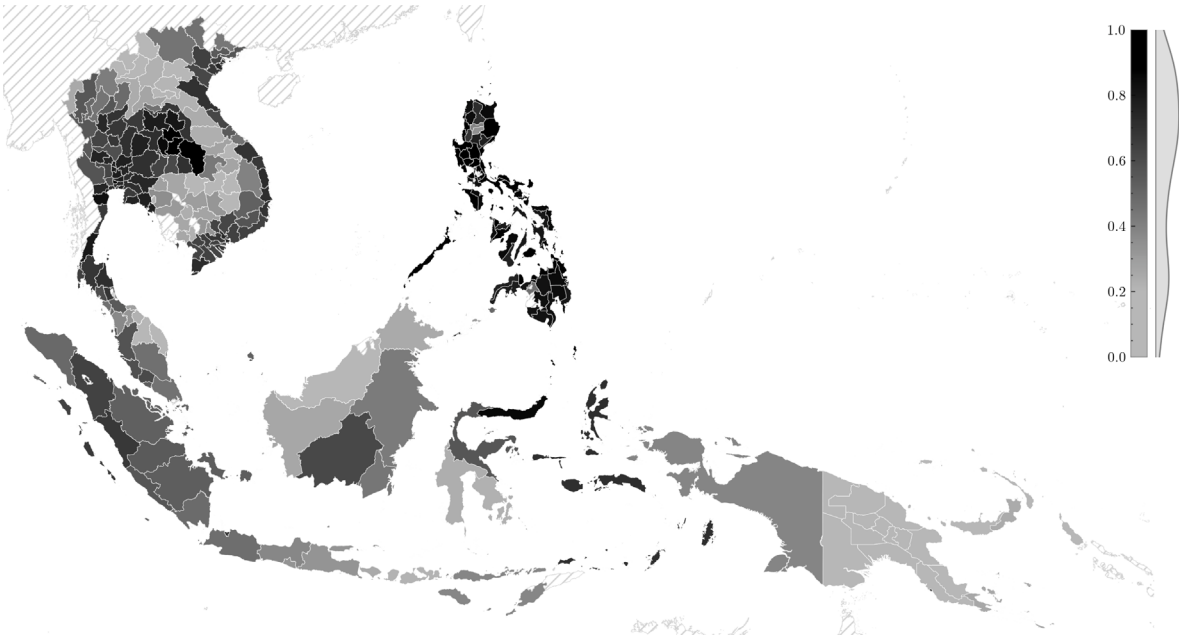

Fig. 2. The share of the 1925 birth cohort with some formal education. *Notes*: Darker is higher. The highest value is 0.97; the lowest value is 0.01; the mean is 0.57.

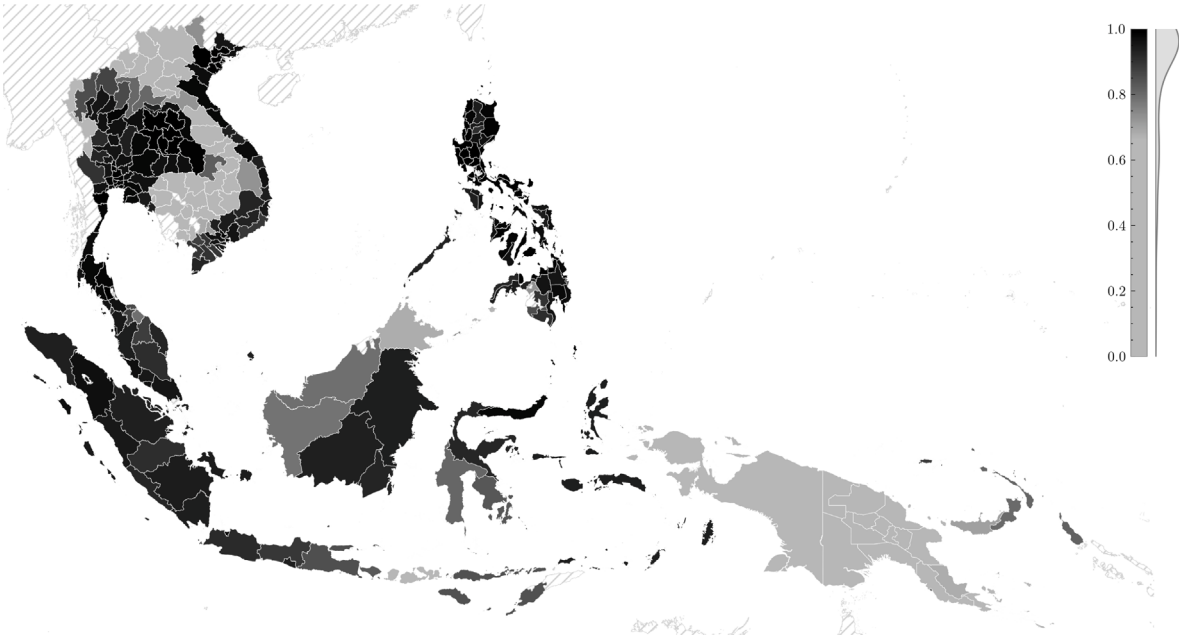

Fig. 3. The share of the 1965 birth cohort with some formal education. *Notes*: Darker is higher. The highest value is 0.99; the lowest value is 0.25; the mean is 0.89.

At least four factors influenced the uneven diffusion of formal education *within* countries. The first factor relates to the emergence of administrative centres. As stated above, the earliest forms of state investment in schooling were geared towards the training of a class of civil servants, and this obviously happened in capital cities. In all eight countries the district of the capital city records attainment shares that are higher than average. Port Moresby in Papua New Guinea has already been singled out as an outlier (43 percentage points above the mean), but for the other capitals (except Bangkok in Thailand) we also obtain differences: i.e. Hanoi (5 pp), Jakarta (Batavia) (22 pp), Kuala Lumpur (16 pp), Manilla (14 pp), Phnom Penh (7 pp) and Vientiane (21 pp). Other major cities such as Medan, Penang or Saigon reveal a similar pattern. It is important to note that in the regressions in Section 4 we will control for the fact that urban centres are positively correlated with school attainment levels as well as with long-run economic development outcomes.

The second factor has to do with features of economic geography. Some areas are more prone to sedentary agriculture or trade than others. Fertile plains with rivers offer better conditions for crop cultivation than deserts or steep mountains. Coastal regions with natural harbors have a commercial advantage over landlocked hinterlands (Henderson et al., 2018). Disease environments and climatic conditions also affect human settlement conditions. For the spread of new ideas and ideologies, including conceptions of 'modernity', such inequalities in economic geography matter. In general, the better connected or richer agricultural areas are more likely to get enmeshed in the

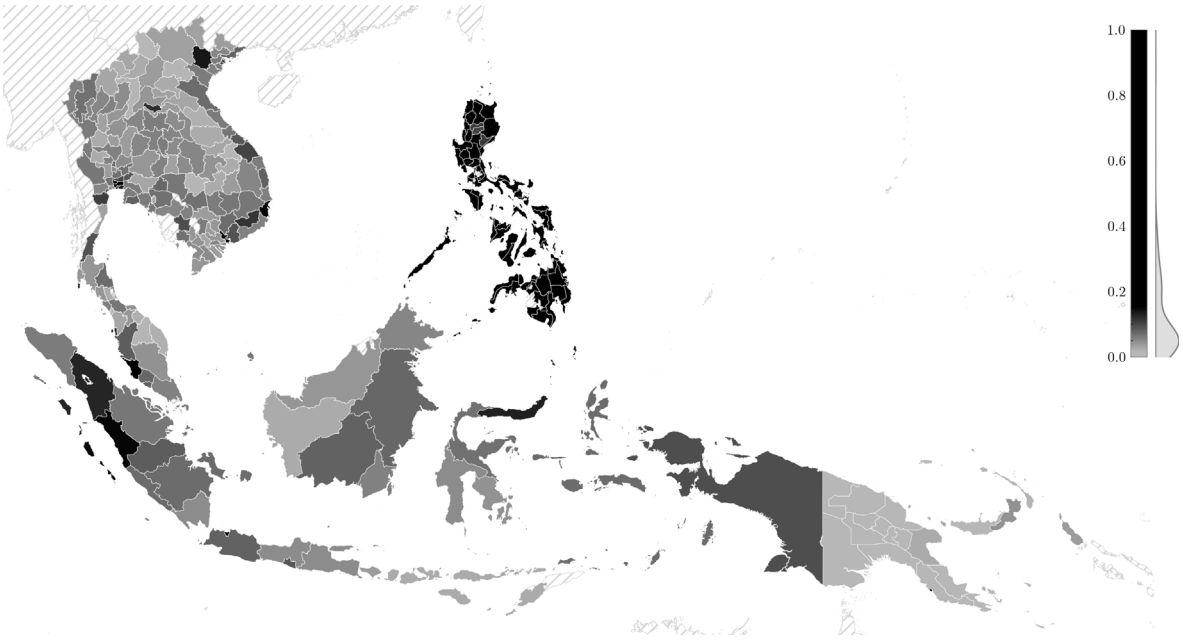

Fig. 4. The share of the 1925 birth cohort with higher education. *Notes*: Darker is higher. The highest value is 0.55; the lowest value is 0.00; the mean is 0.07.

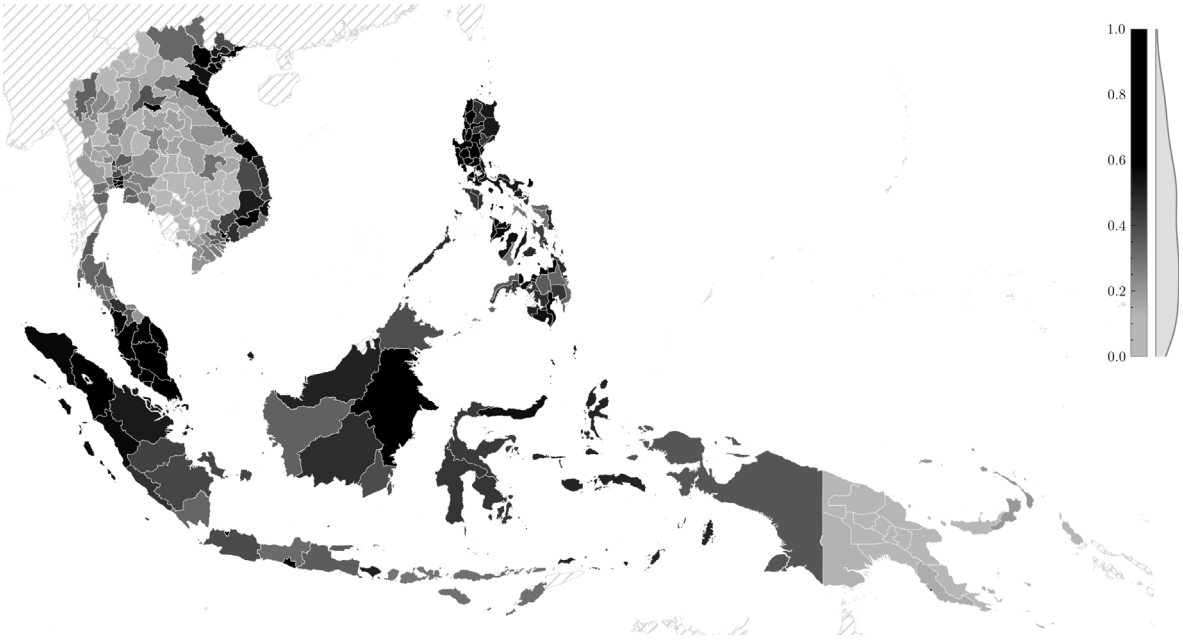

Fig. 5. The share of the 1965 birth cohort with higher education. *Notes*: Darker is higher. The highest value is 0.79; the lowest value is 0.06; the mean is 0.34.

schooling revolution than the more isolated or impoverished regions. Our regressions will include a range of control variables that capture these inequalities in economic geography.

A third factor pertains to local religious and cultural identities. The reception of Western forms of education varied across polities, social classes and ethnic groups. In the Islamic heartlands of Java and Sumatra, Christian missions had a hard time establishing outposts and schools. The Dutch colonial government discouraged and sometimes even prohibited the opening of Christian schools in these areas fearing Muslim resistance and social unrest (Furnivall, 1948, p. 377, Kruithof, 2014). Christian missionaries diverted their attention to the so-called outer islands. As can be seen in Fig. 2, some of these outer areas, such as the Minahassa region in Northern Celebes, had surprisingly high education levels. These were also the result of network effects in missionary location choices: once local reception of Christian missionaries was favourable, churches sought to expand their activities. Closer inspection of the Philippines reveals two areas (Northern central Luzon and the Western part of the Southern Mindanao island) with relatively low education rates. Both of these regions were inhabited by communities who had resisted foreign infiltration (i.e. Hispanization) on cultural and religious grounds for centuries (Phelan, 1959). Tensions between 'local' and 'colonial' suppliers of education, and especially between Islamic and Christian schools, thus led in some countries to regional reversals in educational development.

On the Southeast Asian mainland Buddhist temples were the chief suppliers of schooling before the transition to formal public schooling set in. According to a 1891 British census of upper Burma, 53.2% of the male population aged 5 and over was either literate or in school.

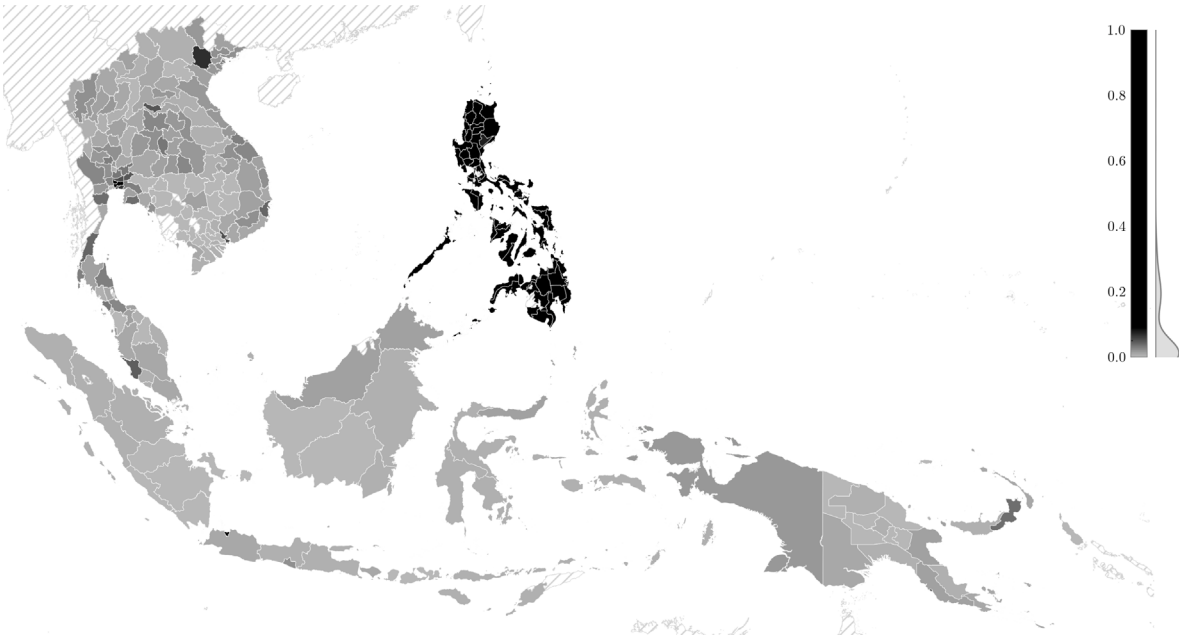

Fig. 6. The share of the 1925 birth cohort with tertiary education. *Notes*: Darker is higher. The highest value is 0.39; the lowest value is 0.00; the mean is 0.03.

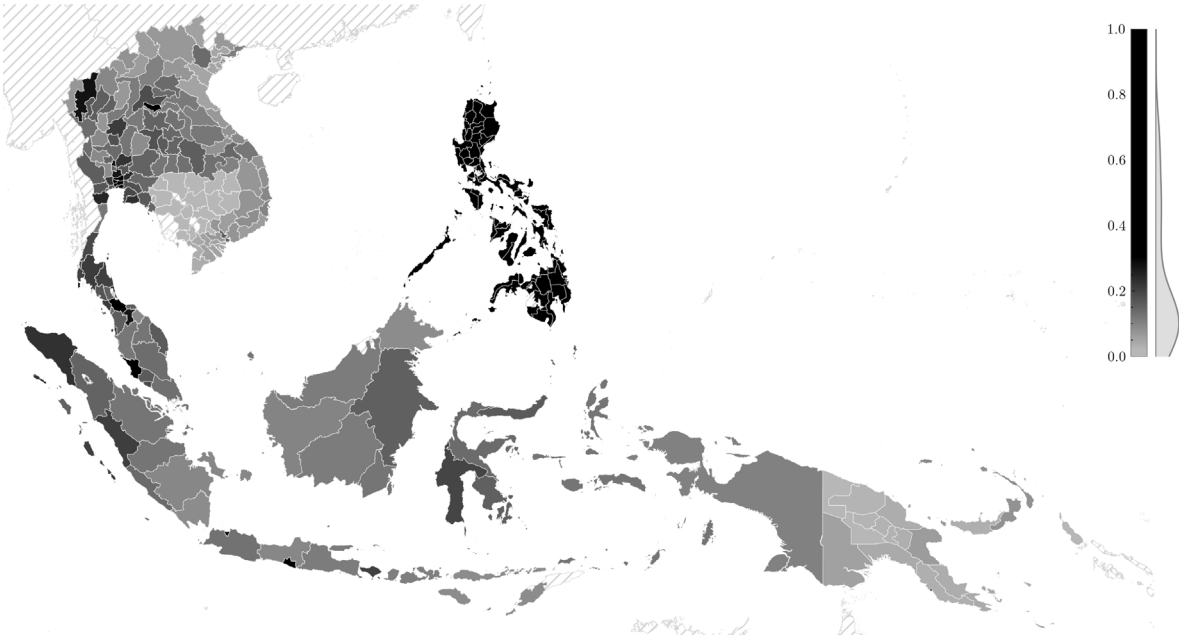

Fig. 7. The share of the 1965 birth cohort with tertiary education. *Notes*: Darker is higher. The highest value is 0.50; the lowest value is 0.02; the mean is 0.12.

This compared to a literacy rate of just 1.5% for female, who were not admitted to temple schools (Lieberman, 2003, p. 189). Data from the *Statistical Yearbook of Thailand* (Central Statistical Office, 1955) offers tentative evidence that the presence of Buddhist monasteries, some of which were established since the 13th century, is associated with higher levels of formal education around the middle of the 20th century. Tables 2.1 and 2.2 show positive associations between the number of Buddhist teachers and monasteries per person and education levels of the 1925BC (Columns (1) and (2)), 1955BC (Columns (3) and (4)) and 1965BC (Columns (5) and (6)). As mass education spread, the relationship between monasteries and the share of people with some education declined as one would expect, but the association with higher levels of education, which expanded later grows even stronger: one additional monastery per person in 1952 is associated with an approximate 9.5% increase in the IHS-transformed share of higher educated people in BC1965. Our regressions include the full set of controls, which we introduce in Sections 3.4 and 4 below.

A fourth factor impacted especially, albeit not exclusively, household *demand* for schooling: the use of child labour. When child labour contributed substantially to household subsistence, or reduced the wage bill of colonial planters, the incentives to send children to school were lower. Parents would either not see the benefits of educating their children, or could not forego the additional income their children brought in. Tobacco planters in Indonesia, who used a lot of child labour, did little to promote the establishment of a school in nearby villages (Breman, 1992). Yet, in other activities, such as large mines or in more capital or land intensive forms of agriculture, labour specialization patterns were more geared towards adult male labour, or simply

Table 2.1
Religion and the share with some education: Thailand.
Sources: See text.

	BC1925		BC1955		BC1965	
	(1)	(2)	(3)	(4)	(5)	(6)
Priests per person, 1952	0.107**		0.0181		0.00879	
	(0.0498)		(0.0226)		(0.0163)	
Monasteries per person, 1952		0.216***		0.0864**		0.0594*
		(0.0697)		(0.0381)		(0.0320)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67	67	67	67	67	67
R-squared	0.77	0.77	0.56	0.60	0.53	0.56

Notes: All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to controls for heteroscedasticity. BC1925, BC1955, and BC1965 refer to the share of the adult population born in respectively the 1920s, 1950s and 1960s, who had received at least some formal schooling at the time the census was taken. The regressions include the full set of controls introduced in Sections 3 and 4. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

Table 2.2
Religion and the share with completed secondary education: Thailand.
Sources: See text.

	BC1925		BC1955		BC1965	
	(1)	(2)	(3)	(4)	(5)	(6)
Priests per person, 1952	−0.00381		0.00114		0.00914	
	(0.00693)		(0.0134)		(0.0194)	
Monasteries per person, 1952		0.0195*		0.0656**		0.128***
		(0.0112)		(0.0279)		(0.0319)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67	67	67	67	67	67
R-squared	0.54	0.55	0.71	0.73	0.74	0.78

Notes: All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. BC1925, BC1955, and BC1965 refer to the share of the adult population that had received at least some formal schooling at the time the census was taken of the 1920s, 1950s and 1960s cohorts respectively. The regressions include the full set of controls introduced in Sections 3 and 4. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

demanded a significant share of skilled workers. In these circumstances, the incentives to send children to a vocational or technical school were higher.

In sum, while these four factors are not exhaustive, they go a long way in helping us understand the inter-regional heterogeneity in education levels of the working age population. This heterogeneity also informs our empirical strategy, in which we employ an expansive set of control variables and exploit the time-distance between early formal education and post-war structural change, rather than relying on an IV that fails to capture the multiplicity of historical and geographic forces that influenced Southeast Asia's schooling revolution.

3. Data

3.1. Human capital formation in the long run

To estimate regional patterns of human capital formation in Southeast Asia we use the anonymized samples of all available censuses for Cambodia, Indonesia, Laos, Malaysia, Papua New Guinea, the Philippines, Thailand and Vietnam provided by IPUMS International (Ruggles et al., 2024). The census data contain basic demographic information, as well as information on religion, occupation, industry, income, work status, education, type of housing, disability status, and household characteristics of ca. 123 million persons (see Appendix A for an overview of the number of individuals in each census). For this study, we extracted information on educational attainment and migration. We aggregated the individual-level data up to the 'province' level (denoted *geolev1* in IPUMS and in corresponding GIS boundary files) using the location of residence of the individuals in the censuses. Since data on place of birth is only available at *geolev1*, the comparison is confined to 277 provinces. The compression of 123 million persons into just 277 regions may appear as a waste of information, but this strategy ensures that also for the thinner (older) birth-cohorts in the lesser populated regions, sufficient coverage is maintained to conduct stable comparisons across time and space. The 277 regions are divided across the eight countries as follows: Cambodia (19); Indonesia (27); Laos (18); Malaysia (13); Papua New Guinea (20); the Philippines (75); Thailand (68); and Vietnam (37).

For almost all individuals IPUMS reports the highest level of schooling completed, including those who received no schooling. This allows us to trace the progress of mass education captured by the share of the working age population with 'some schooling' and to construct two indicators that capture the diffusion of higher education: the share of the population with 'completed secondary education' and the share with 'completed tertiary education'. To project attainment rates back in time, we make use of age information. We create birth-cohorts for six consecutive decades between 1910 and 1969. For each of these birth-cohorts we compute attainment levels: the '1925BC' contains all observations on the schooling achievements of individuals born between 1920 and 1929, the '1935BC' covers all individuals born between 1930 and 1939, and so on. Since most children go to school between 6 and 20 years of age, we restrict our sample to people aged 20 or above. We have removed the very old (90 or above), and made sure our sample is consistent across time and space (see Appendix B for details on harmonization). While we will show regressions for the 1925BC and 1965BC in the main text, Appendix C shows the results for all BCs.

3.2. Survivor bias and other data limitations

Applying a birth-cohort approach to IPUMS micro-data enables researchers not only to trace educational investments back in time, but also to explore chronologies of edcuational development over time without major gaps (cf. McCaa et al. (2015), De Haas and Frankema (2018) and Baten et al. (2021)). The disadvantage of the approach, however, is the time gap that exists between a population's reported attainment level at the time of the census, and the actual schooling

enjoyed in the past. Especially for older generations this can introduce biases due to differential survival rates. As higher educated persons tend to live longer, historical attainment rates may be overestimated. Such survivor biases also tend to be non-linear, they grow stronger as cohorts grow older. Another potential problem is that perceptions of what constitutes 'schooling' (i.e. preschool, adult education, apprenticeships), or 'citizenship' may change over time. Such conceptual shifts can cause measurement inconsistencies across consecutive censuses.

We check the role of survivor bias and other irregularities by taking a closer look at the statistical coherence of successive census reports. We show the data for the categories 'some education' and 'higher education' for each country in Appendix D. As can be seen, survivor bias is a common feature in all the countries in our sample, except for Malaysia, where the pattern appears to be reverse, that is, recorded shares of attainment tend to decline in later censuses instead of rise. A team of scholars affiliated with the Minnesota Population Center which hosts IPUMS, have looked into the cross-census statistical coherence of completed primary schooling shares in 13 Asia-Pacific countries, including 6 of the countries in our sample (Cambodia, Indonesia, Malaysia, Philippines, Thailand, Vietnam; see McCaa et al. (2015)). They too find evidence of survivor bias, but conclude that the statistical coherence is outstanding, with mean differences hoovering between less than one percentage point to a few percentage points at most. McCaa et al. (2015, Table 4 pp. 24) also show that the coherence for Malaysia is more problematic. While we are unable to ascertain why reported attainment shares are declining across the census of 1991 and 2000, we believe that it is related to the exceptionally large influx of international labour migrants from India and China before 1940. In the 1920s, already ca. 50% of the total registered population had a migrant background (Sultan Shah, 2017, pp. 42) and we find that the provinces that attracted most of the migrants reveal the clearest downward trend. It is possible that census-takers' perceptions of the status of the education attained by migrant children has changed (migrant communities established their own vernacular schools) or that larger groups of lower educated migrants have applied for 'citizenship' and became included in later rounds of censuses, thus suppressing the mean in those provinces where they settled.

In order to make sure our results are robust to such biases our main strategy is to take the averages of provincial attainment shares across successive censuses, which will give more weight to earlier censuses with birth-cohorts less affected by mortality. We also performed two additional checks by running all regressions excluding censuses after the year 1999, and by excluding Malaysia. The main results, shown in Tables D.17 and D.18, are robust to these alternative sampling strategies.

Survivor bias may also operate in another direction, namely through higher mortality of the higher educated segments of the population, so that late 20th century censuses are underestimating historical attainment levels. The most infamous case in this regard is the genocide of ca. 25% of the Cambodian population perpetrated by the Khmer Rouge regime under command of Pol Pot in 1976–78. This genocide focused in part on the educated, who were regarded as enemies of agrarian communist ideology (de Walque, 2005). Wearing glasses could already be lethal, as it was associated with intellectualism. The killing of many professionals, including teachers, doctors, and former government officials, shows up in Figs. 7 and 9, and especially when comparing the development of higher education in Cambodia with neighbouring Laos. It is also possible, although we are unaware of any estimates, that the mass killings of communists in Indonesia in 1965–66 disproportionally affected the higher educated parts of the population in some provinces. Regarding the bias introduced by differential survival as a result of political violence, we can only assume that the early death of educated persons has significantly reduced their potential contribution to long-term economic development.

Finally, there is the issue of adult education. As our data is restricted to the recorded attainment levels of all individuals aged 20 to 90, it may be the case that people who enroll in adult schooling programmes report different levels in later censuses. This phenomenon appears in the census data of Vietnam, where we observe a sharp rise in the higher education attainment shares of the 1955BC and 1965BC, which includes people born in the middle of the Vietnam war, when comparing the census of 2009 with 1999. This rise is difficult to ascribe exclusively to survivor bias. As explained above, we take care of this issue by splitting the dataset into a pre-2000 and post-2000 sample, as well the standard inclusion of country fixed effects.

3.3. Measuring economic development at the sub-national level

We follow the recent literature that measures sub-national levels of economic activity by lights at night as observed from satellites (Henderson et al., 2012). GDP per capita, the most important variable in analyses of economic growth and development, is often prone to mismeasurement, especially in developing countries such as Southeast Asia, and not available beyond the national level. Henderson et al. (2018) offer nightlight density data in 2000 and 2012, which are distributed as a grid of pixels of dimension 0.5 arcminutes (approximately 1 square kilometre at the equator) which we have aggregated up to an average for each of the 277 provinces in our sample (see Fig. 8). Our baseline regressions, presented in Section 4, use lights at night in 2000 as the dependent variable. Appendix E presents the same regressions using the 2012 values that show similar results.

3.4. Confounding factors

The regression analyses presented in Section 4 account for a wide range of potentially confounding factors. We extensively control for the geographical characteristics of each region, as well as the region's initial level of development, particularly variations in degrees of urbanization. All of the confounding factors for which we include proxies may have either contributed to Southeast Asia's take-off after 1970 and/or to the development of formal education. Our control variables, including a description of sources, are listed in Appendix F.

Our set of geographical controls includes average temperature, average rainfall, elevation, a dummy for access to the sea, a dummy for having a main river in the region, latitude, longitude, a dummy indicating whether or not a border is shared with a neighbouring country, and the size of the region. As conventional indicators of land quality were too granular, we capture possible effects from land quality using wet- and dry-rice suitability, which is the largest and most important agricultural sector in Southeast Asia, comprising up to 90% of total agricultural production. To capture the possible extent of cash crop production in colonial times, we include proxies for rubber, sugar, tobacco, and coffee suitability. Finally, to control for potential effects from mineral extraction, we include proxies for the presence of coal, oil, and tin.

To control for initial levels of economic development, we calculated the distance from the centroid of each province to the nearest city in 1900 with more than 5000 inhabitants (data from clio-infra.eu). This control variable, together with the size of the region, captures much of the variation in early levels of urbanization. The size of the region is important, as it reflects the specific role that capital cities play in attracting both human capital and economic activity. Capital cities in all eight countries are a region in their own right, and therefore much smaller than all other provinces.

4. Empirical analysis

4.1. Exploring patterns of economic development

How strong is the association between early human capital formation and long-run economic development in Southeast Asia? Our

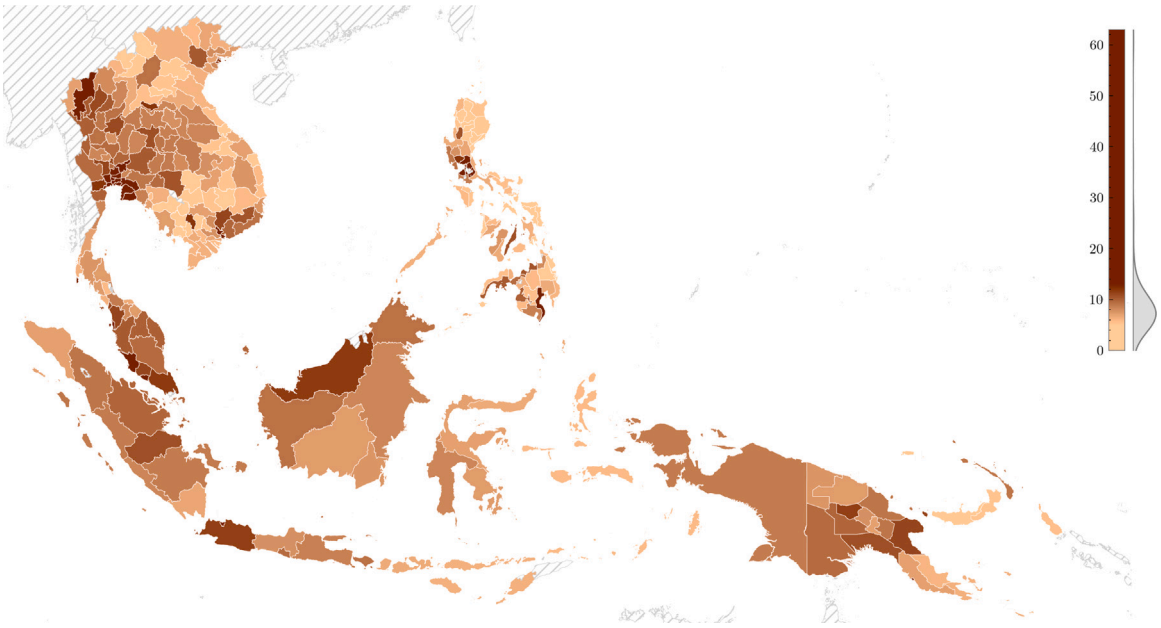

Fig. 8. Lights at night, 2000. *Notes*: Darker is higher. The highest value is 63; the lowest value is 0; the mean is 9.5.

analysis begins with a cross-sectional estimate of the relationship between educational attainment levels and nightlight density using the following equation:

$$\text{Econ Dev}_r = \alpha + \beta_1 \text{HC}_{bc,r} + \text{X}'_r\mu + \Omega_c + \varepsilon_r, \quad (1)$$

where Econ Dev_r is night light intensity in region r in 2000, $\text{HC}_{bc,r}$ is the share of the birth cohort bc residing in region r with a given education level, Ω_c represents country fixed effects, and ε_r is the error term. As discussed in Section 3.4, X'_r includes a set of geographical controls as well as the initial levels of development (the distance to the nearest city with more than 5000 inhabitants in 1900). Appendix G provides summary statistics for all variables used. The baseline regressions use an inverse hyperbolic sine transformation, which can be sensitive to units of measurement when zeros are present in the data (Chen and Roth, 2024; Mullahy and Norton, 2024). To address this, alternative methods, including Poisson Pseudo-Maximum Likelihood models, are provided in Appendix H. Additionally, Appendix I presents results that account for potential spatial autocorrelation in the error terms or autoregressive processes in the outcome variable. The results from these alternative models and adjustments for spatial autocorrelation closely align with those reported in Sections 4.2 and 4.3.

4.2. Higher education matters most

Columns (1), (4), and (7) in Table 4.1 report the baseline version of our model. These columns include the share of the 1925BC with some education (1), completed secondary education (4), and completed tertiary education (7), along with country fixed effects and our proxy for the initial level of economic development, i.e. the distance to the nearest city with more than 5000 inhabitants in 1900. All variables enter the regression with the expected signs and significance levels, and the coefficients on our education variables remain remarkably stable once we add the geographical controls (columns (2), (5), and (8)) and the controls for cash crop suitability and subsoil deposits (columns (3), (6), and (9)).

What is the magnitude of the effects we find? As indicated, the endogeneity inherent to spatial accumulations of human capital makes it tricky to interpret the coefficients as causal. Specifically, as the trade boom under colonial rule increased the demand for educated workers, the observed relationship between human capital and economic development can go both ways (Galor and Weil, 2000). However, the diverse set of controls we use allows us to make a tentative estimation. The size of the effect cannot be obtained directly from the coefficients reported in Table 4.1, as our variables are transformed by the inverse hyperbolic sine (IHS) function, which approximates a log transformation for large values but remains defined for small or zero values. To approximate the magnitude of these effects, we compute marginal effects at the mean of each educational variable, using the results from the fully specified models in columns (3), (6), and (9). A one standard deviation increase in the IHS-transformed share of the population with some formal education (SD: 0.229) is associated with an estimated 16.5% increase in IHS-transformed nighttime lights in 2000. For higher education (SD: 0.072), the corresponding effect is 44.2%, and for tertiary education (SD: 0.048), it is 40.2%. Among the three educational variables considered, higher and tertiary education have the strongest impact on regional development in 2000. General education, while still impactful, has a relatively smaller effect.

Let us now turn to the generation born 40 years later, the 1965BC, noting that the results for all the intermittent birth cohorts are presented in Appendix C. For this younger generation, we observe that the effect of the share of the workforce with some formal education on economic development has turned insignificant altogether (see Table 4.2). This is not surprising, given the fact that the vast majority of this birth cohort received at least some schooling, reducing the variation across regions. For the shares with higher education, the effect of those with completed secondary education on both measures of development remains strong and positive. Specifically, using the coefficients on education in columns (6) and (9), a one standard deviation increase in higher education in 1965 (0.161) is associated with a 40.0% increase in IHS-transformed nighttime lights in 2000. For completed tertiary education (SD: 0.108), this corresponds to a 49.7% increase.

In sum, the regression results show that early investments in higher education are particularly strong predictors of early 21st-century development levels. The next question is whether these results also hold when we adjust for migration across regions. Put differently, to what extent are cross-regional migrants responsible for human capital accumulation, and what happens if we take them out of the equation?

4.3. Exploring the contribution of migration

There exists a large theoretical and empirical literature on rural–urban migration that addresses the question why labour productivity

Table 4.1
Attainment shares of the 1925BC and economic development.
Sources: See text.

	IHS lights at night, 2000								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education, 1925	0.831***	0.885***	0.784***						
	(0.234)	(0.244)	(0.247)						
Higher education, 1925				6.125***	6.287***	6.477***			
				(0.490)	(0.624)	(0.609)			
Tertiary education, 1925							8.340***	7.831***	7.933***
							(1.092)	(1.168)	(1.056)
Pre-development	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Cash crops and mining	No	No	Yes	No	No	Yes	No	No	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277	277	277	277	277	277	277	277	277
R-squared	0.45	0.51	0.53	0.61	0.64	0.65	0.55	0.58	0.60

Notes: All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

Table 4.2
Share of the 1965BC with education and economic development.
Sources: See text.

	IHS lights at night, 2000								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education, 1965	0.371	0.351	0.340						
	(0.417)	(0.513)	(0.507)						
Higher education, 1965				1.994***	2.587***	2.615***			
				(0.302)	(0.386)	(0.406)			
Tertiary education, 1965							4.264***	4.485***	4.619***
							(0.635)	(0.671)	(0.713)
Pre-development	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Cash crops and mining	No	No	Yes	No	No	Yes	No	No	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277	277	277	277	277	277	277	277	277
R-squared	0.43	0.49	0.51	0.51	0.59	0.60	0.52	0.57	0.59

Notes: All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

is higher in urban economies. This literature tends to split up the productivity advantages in three components: (1) Exposure: personal skill-sets develop more rapidly due to the learning effects of larger and more diverse social networks; (2) Alignment: people are confronted with a broader range of occupations and can thus select a job that is better tailored to their personal abilities; (3) Positive sorting of rural–urban migrants as well as negative sorting of migrants who decide to swap the city for the countryside (Glaeser, 1999; Young, 2013; Andersson and Molinder, 2024).

Our data does not directly speak to rural–urban migration as it captures migration between regions, but the question of migrant-selectivity is equally valid. Moreover, our data on inter-regional migration has the advantage that we do not have to rely on a proxy of human capital such as occupational wages as is common in most studies: we directly observe the inequalities in educational attainment between migrants and non-migrants. What we do not know, however, is where exactly our migrants received their education, that is, whether they went to school in the region of origin or destination. Keeping these advantages and limitations in mind, we explore the contribution of migrants to regional human capital stocks and subsequently estimate their impact on regional economic development, by taking the non-migratory (sedentary) part of the workforce as a benchmark. We exclude Vietnam from the analysis at this point, as the censuses of this country do not report the place of birth.[5] We adopt the term 'sedentary' for persons who did not migrate (or had returned to their birth region before the census was taken), but note that many of these people were of course not strictly sedentary, as they still may have moved from the countryside to a nearby city *within* their birth region.

We first show an important stylized fact of migration at the country level: impressive degrees of migrant-selectivity. Fig. 9 shows a bar graph presenting the percentage shares of the working age population of all birth cohorts with completed secondary education per country, as well as the shares of the sedentary and the migratory groups separately. In all Southeast Asian countries inter-regional migration was of sizeable proportions, ranging from 14% of the working age population in Cambodia and Papua New Guinea to 30% in Malaysia (for Indonesia: 15%; Laos: 17%; the Philippines: 25%; and Thailand: 20%). International migrants make up only a small proportion of the total group of migrants (also because only first generation immigrants are registered as such). Part of the relatively high migrant share of Malaysia (30%), however, is explained by the influx of Indian and

[5] In addition, the data for the Philippines is based on the 1990 census, as the other censuses do not include information on place of birth. However, the number of individuals in the 1990 census is sufficiently large to construct reliable education shares: see Appendix A.

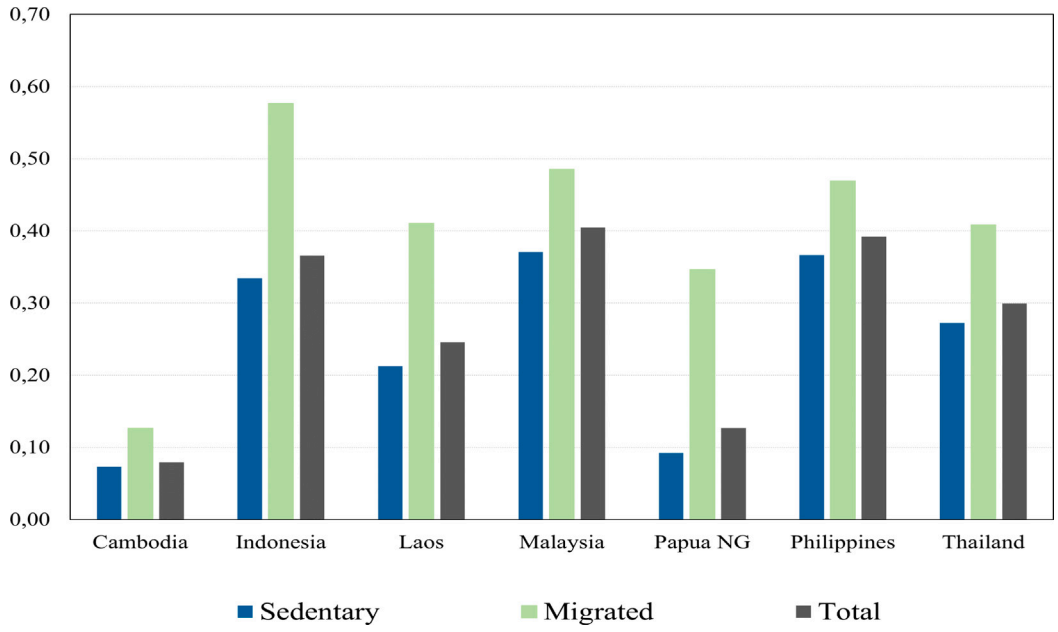

Fig. 9. Share of the sedentary and migratory working age population with completed secondary education in Southeast Asian countries, ca. 2000.
Sources: See text.

Chinese migrants. The considerable share of lesser educated Indians and Chinese may explain why the education gap between migrants and non-migrants is considerably smaller in Malaysia than in neighbouring countries.

Fig. 9 shows that in all societies migrants were considerably more likely to be higher educated. In the lesser developed economies of Papua New Guinea and Laos the education gap is larger than in the more industrialized economies of Thailand, Malaysia and the Philippines. Since migrants constituted a minority, their contribution to the total share of higher educated in the working age population remains confined to c. 2.5 to 3.5 percentage points over and above the levels recorded by the sedentary population. Cambodia stands out for having very low shares of higher educated in both sedentary and migrant populations, which reveals the deep impact that the 1976–78 genocide had on the human capital composition of its labour force. As migrant shares are partly a function of the number and relative size of provinces, we include country fixed effects in the regressions to pick up any of such effects.

What happens if we adjust our regressions by removing all migrants from the sample and estimate the relationship between human capital levels of the non-migrant population and economic development in 2000? Tables 4.3 and 4.4 present the results for the 1925 and 1965BCs, where we have estimated attainment levels on the basis of *place of birth* (denoted 'BP' in variable names) instead of *place of residence*. The regressions show what we would expect based on Fig. 9, namely that the coefficients are smaller than those of the models including migrants (see Tables 4.1 and 4.2), but the effects of educational attainment of the sedentary regional population remain strong and significant.

The next question we ask is whether migrants moved specifically towards the economically more dynamic regions? Fig. 10 shows that they did so indeed. The map suggests that there were two major patterns in Southeast Asian migration. Firstly, all capital cities as well as their wider suburbs (which are sometimes provinces in their own right, e.g. Metro Manila, Nonthaburi or Vientiane Capital) attracted considerable shares of migrants from surrounding regions. Secondly, there are a number of regions which were historically far less urbanized, but which developed important industries or trade hubs during the 20th century. Regions such as Riau on Sumatra (Indonesia), Koh Kong (Cambodia), Davao on Mindanao (Philippines) and Pahang (Malaysia) are all darkly coloured on the map as these regions, once relative backwaters, were populated by large numbers of migrants from adjacent regions. The regressions presented in Table 4.5 show that there exists a strong positive association between regional shares of immigrants and 21st century levels of development, and this correlation is significant for all successive birth-cohorts.

Finally, we are interested in whether migrants' human capital has any additional explanatory power for early 21st century variations in nightlight density. In order to explore this question we run two additional regressions. First, we estimate the impact of higher education attainment shares on nightlights, weighing in the migrant share in the total regional population. The results of this specification are shown in Table 4.6. Second, we take the absolute difference between the share of higher educated of the migrant and the sedentary population of which the results depicted in Table 4.7. The results reveal an important stylized fact of the importance of migration for long-run development: migrants positively selected into the more dynamic regions of Southeast Asia's national economies. Rather than attracting the illiterate and unskilled to emerging labour-intensive urban industries, the 'core' regions benefit from an additional influx of higher educated persons. But the relative size of the migration flow matters. Taking the absolute gap in higher education shares between migrants and non-migrants, the coefficients turn negative. That is, the educational inequalities between the (much) smaller migrant populations in the backward regions and the local sedentary population are larger than the gaps in the 'core' regions. Finally, the results again reveal that migrants with completed tertiary education have larger predictive power for long-run economic development, than migrants with completed secondary education.

5. Conclusion

This paper has explored the deeper roots of Southeast Asia's transition to modern economic growth through the lens of human capital formation and accumulation. Based on an analysis of micro-data of ca. 123 million individuals in 36 censuses that were conducted between 1970 and 2019, we have established a strong and significant relationship between early educational attainment shares and sub-national economic development levels around the year 2000. Using a birth-cohort approach and adopting a large set of historical and geographic control variables, we have argued that the virtuous cycle of feedback loops between economic development and human capital formation can

Table 4.3
Attainment shares of the 1925BC adjusted for migration.
Sources: See text.

	IHS lights at night, 2000								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education BP, 1925	0.671***	0.507**	0.430*						
	(0.234)	(0.243)	(0.243)						
Higher education BP, 1925				5.596***	5.577***	5.898***			
				(0.557)	(0.657)	(0.659)			
Tertiary education BP, 1925							7.308***	6.806***	7.056***
							(0.922)	(1.042)	(1.037)
Pre-development	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Cash crops and mining	No	No	Yes	No	No	Yes	No	No	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241	241	241	241
R-squared	0.45	0.54	0.56	0.56	0.62	0.64	0.52	0.59	0.61

Notes: All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

Table 4.4
Attainment shares of the 1965BC adjusted for migration.
Sources: See text.

	IHS lights at night, 2000								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Education BP, 1965	0.409	0.127	0.101						
	(0.470)	(0.579)	(0.596)						
Higher education BP, 1965				1.896***	2.247***	2.263***			
				(0.378)	(0.457)	(0.530)			
Tertiary education BP, 1965							3.691***	4.031***	4.069***
							(0.670)	(0.672)	(0.790)
Pre-development	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Cash crops and mining	No	No	Yes	No	No	Yes	No	No	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	241	241	241	241	241	241	241	241	241
R-squared	0.44	0.53	0.55	0.50	0.59	0.60	0.51	0.60	0.61

Notes: All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

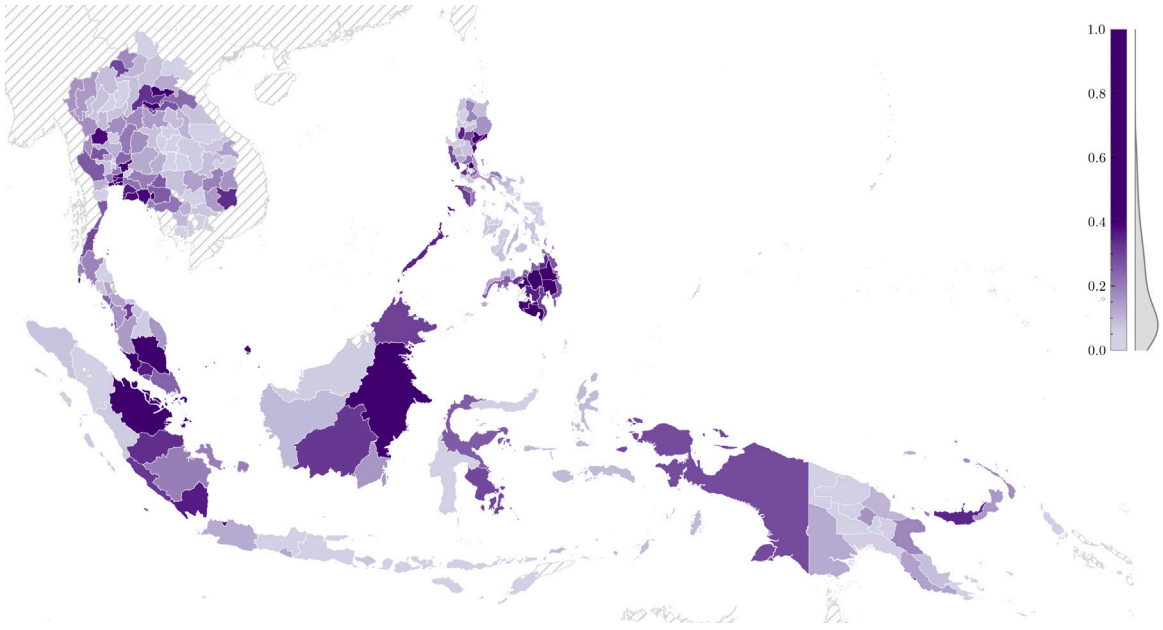

Fig. 10. The share of immigrants, ca. 2000. *Notes*: Purple regions indicate areas with relatively high shares of migrants in ca. 2000. The highest value is 0.98; the lowest value is 0.01; the mean is 0.20.

Table 4.5
Relating the share of immigrants to economic development.
Sources: See text.

	IHS lights at night, 2000		
	(1)	(2)	(3)
Share migrants	1.087***	0.991***	0.945***
	(0.316)	(0.260)	(0.272)
Pre-development	Yes	Yes	Yes
Geography	No	Yes	Yes
Cash crops and mining	No	No	Yes
Country FE	Yes	Yes	Yes
Observations	239	239	239
R-squared	0.50	0.58	0.59

Notes: The share of migrants is calculated using the 2000 census for Malaysia, Papua New Guinea, Thailand and Indonesia; the 1990 census for the Philippines; the 1998 census for Cambodia; and the 2005 census for Laos. All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

Table 4.6
Higher educational attainment of migrants, weighted for migrant share in regional population.
Sources: See text.

	IHS lights at night, 2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Higher education migrants, weighted	3.666***	3.560***	3.460***			
	(0.459)	(0.505)	(0.519)			
Tertiary education migrants, weighted				7.450***	7.164***	7.079***
				(1.012)	(1.061)	(1.071)
Pre-development	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	Yes	No	Yes	Yes
Cash crops and mining	No	No	Yes	No	No	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	239	239	239	239	239	239
R-squared	0.59	0.64	0.65	0.59	0.64	0.66

Notes: The variables are calculated using the 2000 census for Malaysia, Papua New Guinea, Thailand and Indonesia; the 1990 census for the Philippines; the 1998 census for Cambodia; and the 2005 census for Laos. All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

Table 4.7
Absolute difference in higher education.
Sources: See text.

	IHS lights at night, 2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Absolute difference higher education	−0.681**	−0.619**	−0.542**			
	(0.296)	(0.259)	(0.255)			
Absolute difference tertiary education				−1.515***	−1.427***	−1.199**
				(0.571)	(0.482)	(0.506)
Pre-development	Yes	Yes	Yes	Yes	Yes	Yes
Geography	No	Yes	Yes	No	Yes	Yes
Cash crops and mining	No	No	Yes	No	No	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	239	239	239	239	239	239
R-squared	0.45	0.54	0.56	0.46	0.55	0.56

Notes: The variables are calculated using the 2000 census for Malaysia, Papua New Guinea, Thailand and Indonesia; the 1990 census for the Philippines; the 1998 census for Cambodia; and the 2005 census for Laos. All variables are transformed by the inverse hyperbolic sine function, except for dummy variables. Standard errors are robust to control for heteroscedasticity. * Significant at 10%; ** significant at 5%; *** significant at 1% level.

be traced back far into the colonial era. This finding adds evidence to a strand of economic historical literature that argues that deeper roots of the post-1970 take-off have been hidden under the surface of the turmoil of decolonization and civil warfare during the middle decades of the 20th century (1930s–1970s). In fact, an analysis of attainment shares held by successive birth-cohorts in the labour force shows that, with some exceptions, investments in both mass and higher education have continued unabated in this period.

We have also shown that higher education, and especially tertiary education shares, explain regional variation in early 21st century development levels much better than variation in lower levels of education. Compared to the Industrial Revolution in Europe, where mass education did appear to have played a major role in catching-up with Britain, the role of higher education seems to have been of critical importance in the 'late' industrializing economies of Southeast Asia, who operated at a larger distance of the global technology frontier. A final result

of our study is the strong and robust contribution of inter-regional migrants to both human capital accumulation and long-run economic development. Even though the relationship between early education and late development also holds when migrants are excluded, the increasing human mobility which was a hallmark of the globalization wave that set in during the second half of the 19th century, supported the concentration of human capital in regions where the demand for skilled workers grew most.

CRediT authorship contribution statement

Alexandra M. de Pleijt: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ewout Frankema:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2025.103506.

Data availability

Data will be made available on request.

References

Altbach, Philip G., Umakoshi, Toru, 2004. Asian Universities. Historical Perspectives and Contemporary Challenges. The Johns Hopkins University Press, Baltimore.

Amsden, Alice H., 1989. Asia's Next Giant: South Korea and Late Industrialization. Oxford University Press, New York.

Andersson, Jonatan, Molinder, Jakob, 2024. Did Cities Increase Skills During Industrialization? Evidence from Rural-Urban Migration. UPEH Working Paper 2024/12.

Bairoch, Paul, Goertz, Gary, 1986. Factors of urbanisation in the nineteenth century developed countries: A descriptive and econometric analysis. Urban Stud. 23 (4), 285–305.

Barro, Robert J., Lee, Jong-Wha, 2015. Education Matters: Global Schooling Gains from the 19th To the 21st Century. Oxford University Press, Oxford.

Bassino, Jean-Pascal, der Eng, Pierre van, 2021. The First East Asian Economic Miracle: Wages, Living Standards and Foundations of Modern Economic Growth in Southeast Asia, 1880-1938. ANU Centre for Economic History Discussion Paper 2021-05.

Baten, Joerg, de Haas, Michiel, Kempter, Elisabeth, Meier zu Selhausen, Felix, 2021. Educational gender inequality in Sub-Saharan Africa: A long-term perspective. Popul. Dev. Rev. 47 (3), 813–849.

Becker, Sascha, Woessmann, Ludger, Hornung, Erik, 2011. Education and catch-up in the industrial revolution. Am. Econ. J.: Macroecon. 3, 92–126.

Benhabib, Jess, Spiegel, Mark, 2005. Human capital and technology diffusion. In: Aghion, Philippe, Durlauf, Steven (Eds.), Handbook of Economic Growth. Vol. 1, Elsevier, Amsterdam, pp. 935–966.

Berendsen, Bernard, Dietz, Ton, Nordholdt, Henk Schulte, der Veen, Roel van, 2013. Asian Tigers, African Lions: Comparing the Development Performance of Southeast Asia and Africa. Brill, Leiden.

Bevan, David, Collier, Paul, Gunning, Jan Willem, 1999. The Political Economy of Poverty, Equity, and Growth: Nigeria and Indonesia. Oxford University Press, New York.

Birdsall, Nancy, Ross, David, Sabot, Richard, 1997. Education, growth and inequality. In: Birdsall, Nancy, Jasperson, Frederick (Eds.), Pathways To Growth. Comparing East Asia and Latin America. Inter-American Development Bank, Johns Hopkins University Press, Washington D.C., pp. 93–130.

Booth, Anne, 1998. Indonesian Economic Development in the Nineteenth and Twentieth Centuries: A History of Missed Opportunities. Macmillan, London.

Booth, Anne, 2007. Colonial Legacies. Economic and Social Development in East and Southeast Asia. University of Hawai'i Press, Honolulu.

Breman, Jan, 1992. Koelies, Planters En Koloniale Politiek: Het Arbeidsregime Op De Grootlandbouwondernemingen Aan Sumatra's Oostkust in Het Begin Van De Twintigste Eeuw. KITLV Uitgeverij, Leiden.

Brown, Ian, 2013. Burma's Economy in the Twentieth Century. Cambridge University Press, Cambridge.

Central Statistical Office, 1955. Statistical Yearbook of Thailand, vol. 1, Bangkok, pp. 1945–1955.

Chen, Jiafeng, Roth, Jonathan, 2024. Log with zeros? Some problems and solutions. Q. J. Econ. 139 (2), 891–936.

De Haas, Michiel, Frankema, Ewout, 2018. Gender, ethnicity, and unequal opportunity in colonial Uganda: European influences, African realities, and the pitfalls of parish register data. Econ. Hist. Rev. 71 (3), 965–994.

Dell, Melissa, Olken, Benjamin A., 2020. The development effects of the extractive colonial economy: The dutch cultivation system in java. Rev. Econ. Stud. 87 (1), 164–203.

Diebolt, Claude, Chapelain, Charlotte Le, Menard, Audrey Rose, 2021. Neither the elite, nor the mass. The rise of intermediate human capital during the french industrialization process. Cliometrica 15, 167–202.

Frankema, Ewout, 2013. Why was the dutch legacy so poor? Educational development in the Netherlands indies, 1871-1942. Masy. Indones. 39 (2), 307–326.

Frankema, Ewout, Waijenburg, Marlous van, 2023. What about the race between education and technology in the global South? Comparing skill premiums in Colonial Africa and Asia. Econ. Hist. Rev. 76 (3), 941–978.

Fry, Gerald W., 2018. Education in Thailand: An Old Elephant in Search of a New Mahout. Springer Nature Ltd, Singapore.

Furnivall, J.S., 1941. Progress and Welfare in Southeast Asia: A Comparison of Colonial Policy and Practice. Institute of Pacific Relations, New York.

Furnivall, J.S., 1943. Educational Progress in Southeast Asia. Institute of Pacific Relations, New York.

Furnivall, J.S., 1948. Colonial Policy and Practice. A Comparative Study of Burma and Netherlands India. Cambridge University Press, Cambridge.

Galor, Oded, 2011. Unified Growth Theory. Princeton University Press, Princeton.

Galor, Oded, Weil, David, 2000. Population, technology, and growth: from malthusian stagnation to the demographic transition and beyond. Am. Econ. Rev. 90 (4), 806–828.

Geertz, Clifford, 1963. Agricultural Involution: The Process of Ecological Change in Indonesia. University of California Press, Berkeley, CA.

Groningen Growth Development Centre, 2014. 10-Sector database, 2014 release. Available at: https://www.rug.nl/ggdc/structuralchange/previous-sector-database/10-sector-2014?lang=en.

Glaeser, Edward, 1999. Learning in cities. J. Urban Econ. 46 (2), 254–277.

Goldin, Claudia, Katz, Lawrence, 1998. The origins of technology-skill complementarity. Q. J. Econ. 113, 693–732.

Hanushek, Eric, Woessmann, Ludger, 2011. The economics of international differences in educational achievement. In: Hanushek, Eric, Machin, Stephen, Woessmann, Ludger (Eds.), Handbooks in Economics, Volume 3: Economics of Education. Elsevier, Amsterdam, North-Holland, pp. 89–200.

Henderson, Vernon, Squires, Tim, Storeygard, Adam, Weil, David, 2012. Measuring economic growth from outer space. Am. Econ. Rev. 102 (2), 994–1028.

Henderson, Vernon, Squires, Tim, Storeygard, Adam, Weil, David, 2018. The global distribution of economic activity: Nature, history, and the role of trade. Q. J. Econ. 133 (1), 357–406.

Henley, David, 2012. The agrarian roots of industrial growth: Rural development in South-East Asia and sub-Saharan Africa. Dev. Policy Rev. 30 (1), s25–s47.

Henley, David, 2015. Asia-Africa Development Divergence. A Question of Intent. ZedBooks, London.

Hoang, Trung X., Nguyen, Ha, 2023. Multi-generational effects of school access in a developing country: Evidence from a mass education program in Vietnam. Econ. Educ. Rev. 97, 102481.

Huff, Gregg, Angeles, Luis, 2011. Globalization, industrialization and urbanization in pre-world war II southeast Asia. Explor. Econ. Hist. 48, 20–36.

Huff, Gregg, Caggiano, Giovanni, 2007. Globalization, immigration and lewisian elastic labor in pre-world war II southeast Asia. J. Econ. Hist. 67 (1), 33–68.

Hup, Mark, Zwart, Pim de, 2025. Brain Gain in Late Colonial Indonesia: New Evidence on Chinese Migration and Wages. CEPR Discussion Paper 18748.

ILO, 1997. Economically Active Population 1950-2010. International Labour Organization, Geneva.

Ingram, James C., 1971. Economic Change in Thailand 1850-1970. Stanfrod University Press, Stanford, CA.

Johnson, Chalmer, 1982. MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975. Stanford University Press, Stanford, CA.

Katz, Lawrence F., Margo, Robert A., 2014. Technical change and the relative demand for skilled labor: The United States in historical perspective. In: Boustan, L.P., Frydman, C., Margo, R.A. (Eds.), Human Capital in History: The American Record. University of Chicago Press, Chicago, pp. 15–57.

Kaur, Amarjit, 2004. Wage Labour in Southeast Asia Since 1840. Globalisation, the International Division of Labour and Labour Transformations. Palgrave Macmillan, Basingstoke.

Kelly, Gail P., 1977. Colonial schools in Vietnam, 1918 to 1938. In: Proceedings of the Meeting of the French Colonial Historical Society. Vol. 2, pp. 96–106.

Kelly, Morgan, Mokyr, Joel, Gráda, Cormac Ó, 2023. The mechanics of the industrial revolution. J. Political Econ. 131 (1), 59–94.

Kohli, Atul, 2004. State-Directed Development: Political Power and Industrialization in the Global Periphery. Cambridge University Press, Cambridge.

Krueger, Alan B., Lindahl, Mikael, 2001. Education for growth: Why and for whom? J. Econ. Lit. 39 (4), 1101–1136.
Kruithof, M., 2014. 'Shouting in a Desert': Dutch Missionary Encounters with Javanese Islam, 1850–1910 (Ph.D. thesis). Erasmus University Rotterdam, Unpublished.
Laitin, David, Ramachandran, Rajesh, 2022. Linguistic diversity, official language choice and human capital. J. Dev. Econ. 156, 102811.
Lee, Jong-Wha, Lee, Hanol, 2016. Human capital in the long run. J. Dev. Econ. 122, 147–169.
Lieberman, Victor B., 2003. Strange Parallels: Southeast Asia in Global Context, C. 800-1830. Volume 1. Integration on the Mainland. Cambridge University Press, Cambridge.
Lipton, Michael, 1977. Why Poor People Stay Poor: A Study of Urban Bias in World Development. Temple Smith, London.
Mankiw, Gregory N., Romer, David, Weil, David N., 1992. A contribution to the empirics of economic growth. Q. J. Econ. 107, 407–437.
Marks, Daan, 2010. Unity or diversity? On the integration and efficiency of rice markets in Indonesia, c. 1920–2006. Explor. Econ. Hist. 47 (3), 310–324.
McCaa, Robert, Cleveland, Lara, Kelly-Hall, Patricia, Ruggles, Seteven, Sobek, Matthew, 2015. Statistical coherence of primary schooling in IPUMS-international integrated population samples for China, India, Vietnam and ten other Asia-Pacific countries. Chin. J. Sociol. 1 (3), 333–355.
Mokyr, Joel, 2002. The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton University Press, Princeton, NJ.
Mokyr, Joel, 2005. Long-term economic growth and the history of technology. In: Aghion, Philippe, Durlauf, Steven N. (Eds.), The Handbook of Economic Growth. Elsevier, Amsterdam.
Mokyr, Joel, Voth, Hans-Joachim, 2009. Understanding growth in early modern europe. In: Broadberry, Stephen N., O'Rourke, Kevin H. (Eds.), The Cambridge Economic History of Europe. Cambridge University Press.
Mullahy, John, Norton, Edward C., 2024. Why transform y? The pitfalls of transformed regressions with a mass at zero. Oxf. Bull. Econ. Stat. 86 (2), 417–447.
Myrdal, Gunnar, 1968. Asian Drama: An Inquiry Into the Poverty of Nations. Pantheon, New York.
Nayyar, Deepak, 2019. Resurgent Asia: Diversity in Development. Oxford University Press, Oxford, UK.
Nelson, Richard R., Phelps, Edmund S., 1966. Investment in humans, technological diffusion, and economic growth. Am. Econ. Rev. 56 (1/2), 69–75.
O'Rourke, Kevin, Williamson, Jeffrey G., 1999. Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy. The MIT Press, Cambridge, MA.
Phelan, John Leddy, 1959. The Hispanization of the Philippines: Spanish Aims and Filipino Responses, 1565-1700. University of Wisconsin Press, Madison, Wisconsin.
Pritchett, Lant, 2013. Rebirth of Education: Schooling Ain'T Learning. Center for Global Development, Washington.
Ruggles, Steven, Cleveland, Lara, Lovaton, Rodrigo, Sarkar, Sula, Sobek, Matthew, Burk, Derek, Ehrlich, Dan, Heimann, Quinn, Lee, Jane, 2024. Integrated Public Use Microdata Series, International: Version 7.5 [Dataset]. Minneapolis, MN.
Stubbs, Richard, 2018. Rethinking Asia's Economic Miracle. Red Globe Press, London.
Sultan Shah, Nazrin, 2017. Charting the Economy. Early 20th Century Malaya and Contemporary Malaysian Contrasts. Oxford University Press, Oxford.
van Nederveen Meerkerk, Elise, 2017. Challenging the de-industrialization thesis: Gender and indigenous textile production in java under dutch colonial Rule, c. 1830–1920. Econ. Hist. Rev. 70 (4), 1219–1243.
de Walque, Damien, 2005. Selective mortality during the khmer rouge period in cambodia. Population and Development Review.
Woodside, Alexander, 1983. The triumphs and failures of mass education in Vietnam. Pac. Aff. 56 (3), 401–427.
World Bank, 1993. The East Asian Miracle: Economic Growth and Public Policy. Oxford University Press, The World Bank, Washington, D.C..
World Bank, 2024. World development indicators. Available at: https://databank.worldbank.org/source/world-development-indicators.
Wyndham, H.A., 1933. Problems of imperial trusteeship. In: Native Education: Ceylon, Java, Formosa, the Philippines, French Indo-China, and British Malaya. Oxford University Press, London.
Young, Alwyn, 2013. Inequality, the urban-rural gap, and migration. Q. J. Econ. 128 (4), 1727–1785.