



Education Policies and Systems across Modern History: A Global Dataset

Adrián del Río, Carl Henrik Knutsen,
and Philipp M Lutscher

May 2023

Working Paper

SERIES 2023:138

THE VARIETIES OF DEMOCRACY INSTITUTE



UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a unique approach to conceptualization and measurement of democracy. The headquarters – the V-Dem Institute – is based at the University of Gothenburg with 20 staff. The project includes a worldwide team with 5 Principal Investigators, 22 Project Managers, 33 Regional Managers, 134 Country Coordinators, Research Assistants, and almost 4,000 Country Experts. The V-Dem project is one of the largest ever social science research-oriented data collection programs.

Please address comments and/or queries to:

V-Dem Institute
Department of Political Science
University of Gothenburg
Sprängkullsgatan 19, Box 711
405 30 Gothenburg
Sweden
E-mail: contact@v-dem.net

V-Dem Working Papers are available in electronic format at <https://www.v-dem.net>

Copyright ©2023 by authors. All rights reserved.

Education Policies and Systems across Modern History: A Global Dataset ^{*†}

Adrián del Río¹, Carl Henrik Knutsen², and Philipp M Lutscher²

¹Berlin Social Science Centre & Centre for East European and
International Studies

²University of Oslo

April 2023

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 863486). In addition to the authors, Sirianne Dahlum, Felix Haass, and Jonas Schmid, all part of the wider project team at the University of Oslo, were responsible for developing the questions and codebook. We are also grateful to members of the DEMED project team, led by Anja Neundorf at Glasgow University, for very valuable inputs to the codebook. We are especially grateful to our team of excellent Research Assistants, Johanne Heitman Akerø, Hannes Bräuer, Nimo Fuad Ali Hassan, and Anna Nybu for their invaluable assistance in coding education laws and policies worldwide. We thank Ben Ansell, Patricia Bromley, Jane Gingrich, Johannes Lindvall, Agustina Paglayan, Antoni Verger, and Matthias vom Hau for feedback on earlier drafts of this paper, and numerous country experts for helping us to identify data sources and respond to our questions. For access to the data, please contact the authors.

[†]**Corresponding author e-mail:** adrian.delriorodriguez@wzb.eu

Abstract

We introduce a global dataset on education policies and systems across modern history (EPSM), which includes measures on compulsory education, ideological guidance and content of education, autonomy or political control of education institutions, and teacher training. EPSM covers 156 countries with populations exceeding 1 million people and time series extend from 1789 to the present. EPSM opens up for studying several questions concerning political control and the politicized nature of education systems. In addition to describing the measures, we detail how the data were collected and discuss validity and reliability issues. Thereafter, we describe historical trends in compulsory primary and secondary education, civics and ideology education, state funding and operation of education institutions, and laws regulating the training of teachers. Finally, we illustrate how our data can be used to address key questions about education and politics, replicating and extending recent analyses on the (complex) relationship between education and democratization.

Keywords: Education; Education policy; Regime type; Democratization

Wordcount: Approx. 7,400 words (excl. frontpage and Online Appendices)

Introduction

Education plays several fundamental roles in shaping the lives of citizens and societies worldwide, and it has done so for centuries. For instance, scholars concerned with “human development” typically point to education, alongside health, as a key factor in shaping the capabilities, autonomy, and life prospects of individuals (Sen, 1999). Economists highlight how “human capital” – a function of the workforce’s education level – is an important determinant of both medium- (Mankiw, Romer and Weil, 1992) and long-term (Romer, 1990) economic growth. Further, social scientists address how education shapes core values and beliefs as well as everyday interactions between individuals (e.g., Lipset, 1959). Finally, education may influence wider political processes and developments such as prospects for revolutions (Dahlum, 2019), civil wars (Thyne, 2006), democratization (Murtin and Wacziarg, 2014), and nation- and state-building (Paglayan, 2022). Insofar as education affects (at least some of) these important outcomes, the policies that influence the provision and content of education are important.

In this paper, we introduce a new cross-national dataset – Education Policies and Systems across Modern History (EPSM) – which contains 21 variables pertaining to the a) existence and nature of compulsory education, b) ideological guidance and content of education, c) autonomy or political control of education institutions, and d) training of teachers. The dataset thus centers on aspects of political control and predominantly contains variables that capture *de jure* characteristics of education policies and systems. The coding, which has extended over several years, has been done by four Research Assistants and one co-author, relying on both primary- (e.g., education laws and decrees) and secondary sources (e.g., scholarly works, education reports and newspapers). The

geographical and temporal scope of EPSM is unprecedented among datasets on education systems and policies. EPSM covers 156 polities with a current population exceeding one million inhabitants, and time series go as far back as the late 18th century for several countries.

While education most certainly matters for important individual and societal outcomes, the exact shape and strength of its effects often remain unclear or contested. One reason pertains to the inherent difficulty of drawing inferences about the (causes and) effects of several education policies and outcomes, as researchers inevitably have to rely on observable data. While the data in EPSM cannot resolve all such causal inference issues,¹ its detailed information and extensive scope can help limit some such problems. The education policy measures that have been available to researchers, so far, typically have limited spatial and temporal coverage, or they are proxy measures. For example, popular, but fairly distal, proxies of (different) education policies and system characteristics are education *outcomes* such as average years of schooling or gross enrollment rates. EPSM includes much more specific measures closer to many of the concepts of theoretical interest in debates on how education policies relate to, e.g., democratization, societal polarization, nation-building, state-building, inequality, or economic growth. Hence, we hope that our new data will enable researchers to arrive at better-founded empirical answers to various research questions of substantive interest.

In the following, we briefly describe existing cross-national education datasets.²

Next, we describe the coverage and contents of EPSM. In extension, we discuss how

¹Important keywords are effect heterogeneity (e.g., an education policy may have quite different impacts on workforce productivity in two different historical or geographical contexts) as well as selection biases and reciprocal causality (e.g., a particular education policy may increase nationalist sentiment or regime loyalty, but these latter factors may also explain whether and how the policy is implemented).

²We thank ANONYMIZED for allowing us access to their recently compiled overview of existing education datasets. This overview was invaluable when writing up this section.

these data were collected, the sources used, and reliability and validity characteristics. After a validation exercise, we present a series of descriptive statistics and discuss several noteworthy patterns in our new data. In the penultimate section, we build on recent work on education and democratization by Paglayan (2022), and present applications of our data on this topic. When doing so, we corroborate Paglayan’s core findings on a much larger sample and, in extension, report several new findings. In the concluding section, we sum up the main features of EPSM and key empirical findings from this paper before discussing several avenues of future research that EPSM opens up.

Existing education datasets

Most cross-national studies on the political-institutional or economic-developmental correlates of education have relied on fairly aggregated measures that capture population-wide educational outcomes. These measures typically zoom in on the “quantity” of education provided (see Dahlum, 2017). Examples include the World Bank’s measures of school enrollment rates for 227 countries across 1970-2020. Another example, which is used in many panel analyses of the causes or effects of education, is the enrollment rate and average years of schooling measures in the Barro and Lee dataset (Barro and Lee, 2013; Lee and Lee, 2016). The long time-series version of this dataset extends across 1820-2010, with measurements every fifth year, and it covers 111 countries. Yet, the vast majority of these historical time series are extrapolated, and only nine countries have non-extrapolated data from before 1870 (Paglayan, 2021, 186).

Cross-country studies drawing on outcomes that indicate the “quality” or other more specific characteristics of the education provided are fewer, reflecting the limited number of measures of such characteristics with good cross-country and temporal cov-

erage. One exception is the education quality dataset by Angrist, Patrinos and Schlotter (2013), which covers 128 countries and includes measurements from 1965-2010 in five-year increments. These data are constructed from students' scores on different international tests and – given the paucity of such tests and lack of tests taken in all countries – involve several non-trivial assumptions for calculating and imputing scores.

Other datasets contain narrower proxies of education quality or more direct measures of education policies but are more limited in temporal coverage. One example is Talis (Ainley and Carstens, 2018), which is the OECD's survey on teaching and learning. The survey is answered by teachers in different countries and measures various factors pertaining to, e.g., teaching, school climate, and professional development. However, the survey is only conducted in 2008, 2013, and 2018 and covered 46 countries in the latter year. The OECD is, more generally, a major provider of recent education statistics, and in their Education at a Glance reports, the OECD compares and contrasts countries on numerous education policies and systems characteristics. Examples are teacher salaries, students per teachers ratios, and characteristics of principals. Yet, the OECD's data cover only the most recent years and comprise OECD countries alongside a few selected "partner countries." While providing rich sources of information about cross-national variation in current education systems and policies, these data can unfortunately not be used to answer questions pertaining to the historical development of education systems or – given the short time series – leverage the time dimension in panel analysis on correlates of education policies.

There are, however, some data collection efforts that have produced more specific measures and extend back to the 19th century. These efforts then focus on fewer countries and have a more limited historical range than our EPSM dataset. One notable example is Ansell and Lindvall (2013). They code primary education systems back

to 1870, but the time series end in 1939, and 19 primarily Western countries are included. Another notable example is Paglayan (2021), who, in addition to enrollment rates, codes different variables pertaining to the timing of the first state interventions in education systems across 33 European and Latin American countries. These variables concern “the year when central governments began to (i) fund primary schools, (ii) manage them, (iii) establish a mandatory curriculum for all primary schools, (iv) establish certification requirements for primary school teachers, (v) train prospective teachers, (vi) mandate local authorities to provide universal access to schooling, (vii) mandate free provision for the poor, and (viii) establish compulsory primary education” (p.186). In section 4, we use Paglayan’s measures for validation of our variables that are intended to measure similar aspects.

Table 1. Overview of a selection of existing datasets on education policies and outcomes

Dataset	Type variables	Countries	Years
World Bank	Education outcomes (quantity)	227	1970-2020
Barro and Lee	Education outcomes (quantity)	111	1820-2010
Angrist et al.	Education outcomes (quality)	128	1965-2010
Talis (OECD)	Education policies	46	2008-2018
Ansell and Lindvall	Education policies	19	1870-1939
Paglayan	First education policies and outcomes (quantity)	33	1720 (first policy)/1810-2010
WERD (Bromley et al.)	Education reforms	189	1970-2020
V-Indoc (Neundorf et al.)	Education systems and policies (de facto)	160	1945-2021
EPSM	Education systems and policies (de jure)	156	1789-2020

Finally, several recently published or ongoing education data collection efforts are complementary to ours. We will highlight two very recently published, major initiatives with global coverage that provide quite different (and thus supplementary) measures to those in EPSM.³ First, the World Education Reform Database (WERD) by Bromley et al. (2023) has education reforms as units, covering 10,253 reforms – implemented after 1970 in 189 countries – of different kinds. In contrast to narrower and

³More education-related datasets are listed in Online Appendix A of Neundorf et al. (2023).

more structured coding efforts on specific education system aspects (as in EPSM), the main informational content in WERD is an open text (from public reports/sources) variable describing the reform and reform year and title. Neundorf et al. (2023) also provide an impressive education dataset, V-Indoc, which covers 160 countries and extends from 1945-2021. V-Indoc is also complementary to EPSM although in a different manner than WERD. Indeed, V-Indoc contains several similarly-sounding variables to those in EPSM centering, e.g., on the education system's indoctrination potential and content, centralized control over the system, and politicization of the curriculum. The several corresponding variables reflect intentional coordination between the teams behind V-Indoc and EPSM, which, in combination, enable analyses of de jure vs. de facto correspondences or gaps in education systems. Whereas EPSM mainly registers education laws, V-Indoc aims at capturing practices on the ground (e.g., in the classroom) and therefore uses country experts to gauge how indoctrination and political control over the education system operate, in practice. These differences in content and data collection methods also imply that V-Indoc and EPSM have different validity and reliability characteristics (cf. the discussions in Neundorf et al.'s section 5 with those in our section 4 below). And, given the difficulty of assessing classroom practices back in time (relative to education laws), V-Indoc has much shorter time series than EPSM.

Table 1 presents a condensed overview of the discussed education datasets, summarizing their type of content and coverage. The bottom row displays the same information for EPSM, which we will now turn to.

Data collection process, coverage, and contents

EPSM measures de jure characteristics of education systems and policies. We have coded four dimensions related to *compulsory education*, *ideological content*, *school autonomy*, and *teacher training*. Appendix F contains the codebook. Many of the 21 items included in the codebook are multi-category questions, often nominal ones. This feature, and the possibility to build more comprehensive indices, for example for the four dimensions, means that it is possible to create more variables of theoretical interest than the 21 variables listed in the codebook.

The first set of variables identifies whether compulsory education exists, whether compulsory education is free, years of compulsory education, and groups exempted from compulsory education (e.g., based on gender, ethnicity, region, or occupation), if any. The second set pertains to the ideological profile of standalone civics courses for primary, secondary, and higher education. The measures include laws mandating civics or ideological training at each level, the ideological profile of such training (e.g., nationalist or regime-specific ideology), mandatory military training in schools, laws that ban particular subject topics, and mandated books in school curricula written by former or current leaders. Regarding the third dimension – school autonomy – seven items capture whether government departments of education exist, the level at which curricula are determined, and the entities operating and funding primary and secondary schools and universities. We also coded which entity has central authority over the appointments of principals for primary and secondary schools. Finally, we coded formal teacher training, focusing on teacher training presence, the entity operating this training, and the extent of ideological requirements to become a teacher.

To code these 21 items, we draw heavily on country-specific laws and mandates as well as government reports. We also draw on books, articles, and Ph.D. dissertations, especially from history of education, sociology, and development studies, whenever such sources are available and published in Spanish, English, Portuguese, Russian, Italian, French, Norwegian, Danish, or Swedish. When language limitations or uncertainty about the coding decision existed, we carried out consultations with country experts. Justification of coding decisions, data sources, and uncertainty are recorded for each indicator to enhance reliability and replicability.

Concerning the coding process, one co-author, who coded 64 countries himself, trained four research assistants (RAs) to identify relevant data sources, indicate uncertainty, and differentiate between *de jure* and *de facto* aspects of education policies.⁴ As a result of this training and accumulated coding experiences, we produced an extensive “Rules-of-thumb for coding” document (see Appendix G). This document proved helpful for effective decision-making – especially when education laws were unavailable or data sources were unclear – ensuring that all coders operated by the same explicit rules of thumb rather than different implicit ones.

Coders were assigned to countries primarily based on language expertise. They first collected and examined studies on the history of education to obtain a general overview of key laws and mandates, their background, and the country’s history. Subsequently, they collected and systematized all the education laws mentioned through government websites or law repositories and recorded the 21 variables. Especially before

⁴To exemplify training on the latter distinction, some coders, when starting, assumed that *de jure* ideological training was present when one of the education law’s goals was cultivating citizenship and democratic values among children. The training clarified that observing this goal does not imply that standalone courses with ideological training are legally recognized and should be counted, absent other information. Sometimes the coders only found traces of *de facto* ideological training. On other occasions, standalone citizenship education courses were not strictly mandatory, as parents could choose between, e.g., religion or civics classes for their children.

1900, these data sources were, unfortunately, often unavailable. For such cases (i.e., for the early part of the time series), we relied mainly on scholarly studies. This qualitative information, alongside specific references, is recorded in our dataset to enhance transparency and replicability. On average, we have about 26 sources (e.g., studies, laws, and links with complementary information) per country.

While assessing our data sources, we adopted three general procedures to improve coding accuracy. First, we discarded studies that make statements about a government's indoctrination efforts without explicitly mentioning through which courses or specific mandates indoctrination takes place. Similarly, we do not base coding decisions on studies making imprecise statements, for instance, about the timing and duration of education laws or policies (e.g., "law X came about the mid-19th century and exists to the present day"). In such cases, we used clues provided by these studies to look for further and more specific documentation.⁵ Second, all countries were independently checked by an extra person, and all data sources were cross-checked with studies in the history of education, typically two studies per law. When notable divergences in coding decisions existed or laws were unclear, coders met to discuss these cases, make collective coding decisions, and develop new rules of thumb to handle future similar cases. Third, we consulted with country experts when the precision or quality of the data sources was insufficiently high to ensure reliable and valid coding decisions. Altogether, EPSM took around 2800 hours to code, typically 16-19 hours per country (excluding time consulting experts and team coordination). Despite these efforts, we still had insufficient

⁵Education policies during the German Empire (1871-1918) illustrate this point. We found that education was aimed at "instilling loyalty to the Kaiser, indoctrination that monarchy is the best form of government, creating loyal servants and soldiers" (Theis, 2021). Yet, this information is insufficient for coding the relevant ideological content questions since we need to link these goals to standalone civics- or ideological training courses. However, we did, ultimately, find evidence of this link through the teaching of religion and local history/social studies (Frank-Michael Kuhlemann, 1991). Thus, we characterized the profile of Imperial German ideological training as religious and regime-specific.

information for coding some country-year-question cells. And, information was typically scarcer for some education-system characteristics; thus, while generally low, missingness varies across variables.⁶

EPSM contains 156 polities with a current population exceeding one million inhabitants. The coded polities are chiefly independent states, but for many polities, we also cover long periods when they were semi-independent states (e.g., pre-1905 Norway) or colonies (e.g., pre-1960 Senegal). Although we collected information from 1789, time series coverage varies as we only code country-year observations included in V-Dem (Coppedge and Ziblatt, 2022*b,a*). Figure 1 shows the number of polities coded by year. Roughly 50 countries have time series from the late 18th century and ten additional countries from about 1830 until the turn of the century. Following the V-Dem time series (for details, see Coppedge et al., 2022), the majority of the remaining countries are coded from 1900 onwards. About half are African and Asian former colonies, whereas the other half are new states, mostly emerging around 1990 with the breakdown of Yugoslavia and the Soviet Union.

Validation

Despite the above-described measures taken to ensure high reliability and validity, coding efforts such as ours inevitably face challenges, e.g., because having multiple sources leaves room for interpretation. To further address such concerns, we cross-validate the most comparable items from EPSM against similar measures from Paglayan's (2021) re-

⁶For example, teacher training presence, the content of civics education, and funding for primary schools have 0.03% missing values. In contrast, coders had a harder time identifying information on military education in the curricula, leading to 10.07% missing.

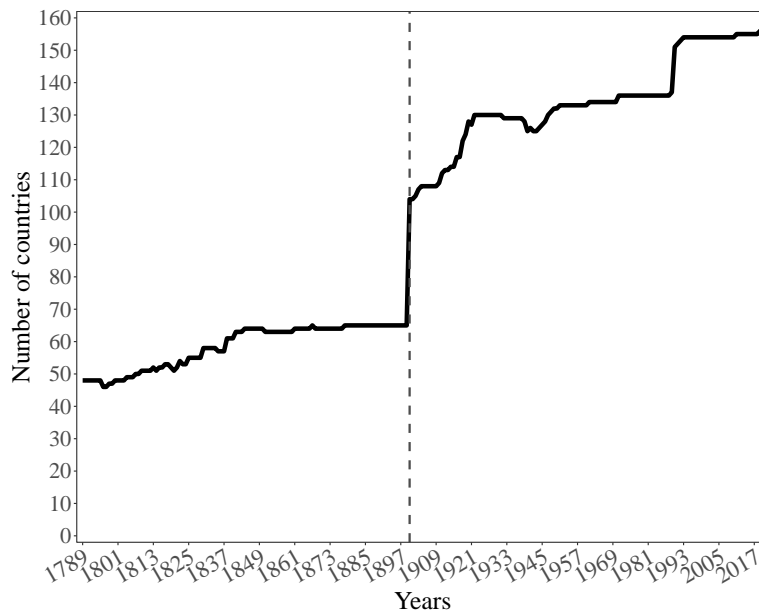


Figure 1. Data coverage, by year, for our measure on compulsory education

cent dataset on initial state interventions in primary education for 32 of the 33 European and Latin American countries that Paglayan covered.⁷

The four charts in Figure 2 compare the coded years for when central governments began to, respectively, (1) establish mandatory education for all primary schools, (2) provide a mandatory curriculum, (3) mandate free education, and (4) train prospective teachers. The dots represent when EPSM identifies a state intervention in primary education, and the circles depict Paglayan’s data points. When circles surround dots, convergence is achieved between the two datasets. This is the norm; the upper-left panel of Figure 2 shows that both datasets identify similar first years for government implementing compulsory primary education in 28 of 32 countries.

The orange squares in Figure 2’s upper-left panel highlight another innovative aspect of EPSM: we also measure when some groups are exempted from compulsory

⁷Trinidad and Tobago (current population < 1 million) is currently not included in EPSM. Going forward, we will expand EPSM by including smaller countries.



Figure 2. Comparing indicators of education from EPSM and Paglayan (2021)

primary education, explicitly or due to the demanding requirements of establishing schools. The relative locations of orange squares and blue dots show that such compulsory education with exemptions for at least one group consistently precedes compulsory primary education for everyone. This new information on (early) compulsory education laws with exemptions may add nuance to descriptive analyses of the introduction of compulsory education globally. It also opens up for investigating the causes and consequences of different types of compulsory education and historical group-based discrimination.

Among the four cases where the datasets diverge, Sweden – coded in 1842/1882 by Paglayan/EPsm – and Bolivia – coded in 1908/1931 by Paglayan/EPsm – are instructive. They illustrate how divergences typically reflect small differences in coding rules and assumptions rather than measurement errors in any of the two datasets. In both cases, Paglayan registers the year of a law that made it mandatory to establish at least one school in medium- and highly-populated towns and that contained financial requirements. According to our coding rules, however, the specific requirements contained in these laws are sufficient for registering that some groups are exempt from compulsory education for geographical reasons (i.e., there is not compulsory education for all). We also find no explicit mention of mandatory schooling or reinforcing practices (e.g., if children are not sent to school, parents must pay a fee or be prosecuted) in these laws. For Bolivia, we note that the native population was included in the education system for the first time only in 1931 (though this was reverted again in 1940). In Sweden, children were exempt from primary compulsory education (Folkskola or ambulatory teachers) if the parents decided to educate children at home or could not attend school due to the absence of routes and other related geographical reasons (Westberg, 2017; Petter-

son, 1991). Thus, de jure compulsory education for everyone first occurred in Sweden in 1882.

A similar high degree of coding convergence is apparent when comparing scores for whether and when governments established a centralized curriculum (upper-right panel Figure 2). In all diverging cases, our *centralized* curriculum coding precedes Paglayan's *state* curriculum coding. This predominantly reflects that many states shared the design and implementation of courses with sub-national governments or other actors (e.g., the church). An exception is Jamaica, where only EPSM identifies that the state regulates the school curriculum (via Paragraph 43(1) of the 1965 education act).

We observe more divergent cases in the two bottom panels, although convergence is still the norm. Specifically, the first year of free education is coded differently in nine cases (bottom-left) and there are eleven cases of disagreement for state-led teacher training (bottom-right). These differences mainly stem from two assumptions that differ between the two data collection efforts. First, EPSM codes whether free education exists for everyone, while Paglayan codes whether the state provides grants for the poor to enroll in primary schools. Note that EPSM does not register the existence of free education when we find evidence of school fees to be paid for one or more groups of students or that grants are provided after state-led competence examination. Second, EPSM explicitly assumes that teacher training questions refer to compulsory education at all levels (and not just for primary schooling). EPSM also codes state-led teacher training (or any other source of training) as missing when there is evidence that more than 50% of teachers do not need a formal degree or training to obtain a job. For example, Costa Rica followed the so-called Lancaster method for acquiring new teachers,⁸ and

⁸With this method, the older or better students (who were later employed) taught the younger pupils.

it was not until the 1869 decree (Decreto Ejecutivo n.º 69, Reglamento de Instrucción Pública) that the state invested in creating schools for training teachers.

To sum up the cross-validation exercise, there is a high degree of convergence between Paglayan’s (2021) and our coding of the initiation of four specific education policies across European and Latin American countries. And, when codings do diverge, we have been able to track down the source of divergence, which are typically minor differences in concepts applied in the two datasets (e.g., “centralized curriculum” vs. “state curriculum”) or differences in particular assumptions (e.g., our added assumption that education is not free, regardless of the text of the law, if there is evidence of categories of students paying school fees or receiving grants after examinations). Hence, the divergences are substantively interesting and seemingly not an indication of low reliability or validity in any of the two datasets.

Historical trends in education systems: Descriptive statistics

This section illustrates the many intriguing real-world patterns that can be identified by using the wealth of information contained within EPSM. Our goal is not to offer detailed discussions and interpretations of particular cross-country patterns or historical trends. Instead, we display and briefly describe several trends across the four dimensions of education policies and systems that we measure.

We start by considering global trends in compulsory education after 1789. Figure 3 contains three plots displaying, by year, the number of countries with compulsory education and with free compulsory education, alongside average years of free compulsory education. The two former plots show a gradual but dramatic expansion – start-

ing from around 1855 and continuing to today – in compulsory education for all (i.e., without exceptions based on gender, class, ethnicity, etc.) as well as *free* compulsory education, globally. While not perfectly correlated, the two plots indicate that, historically, the spread of compulsory education went closely together with the spread of free education for everyone.

There are almost no countries without compulsory education today. And, as of 2020, only nine countries in our sample still lack free compulsory education. Moreover, only three countries combine compulsory education with exceptions for particular categories of students. This was not the case historically. Until 1870, more countries had compulsory education combined with such exceptions (often for girls and remote rural areas) than compulsory education without exceptions, and the former number remained high (around 25% of countries) until World War II, after which it declined gradually.

Regarding average years of compulsory education, we observe a fairly steady increase from an average of two years around 1855 to eight years in 2010.⁹ The small dip around 1900 reflects a sample expansion and, specifically, the inclusion of about 50 former Asian and African colonies, which had comparatively few years of compulsory education at the time. In Appendix A.1, we plot the compulsory education histories of all countries; one observation is that education systems across the world were relatively homogeneous in the decades right after the French Revolution (few countries with compulsory education) and have turned homogeneous again (almost all countries have compulsory education for all) after the 1990s. The intermediate period is characterized by a mix of systems.

⁹Our measure of years of compulsory education is conservative, coding the group that received the fewest years. For example, if children enrolled in rural schools are obliged to four years of schooling and those in urban schools to six years, we code four years.

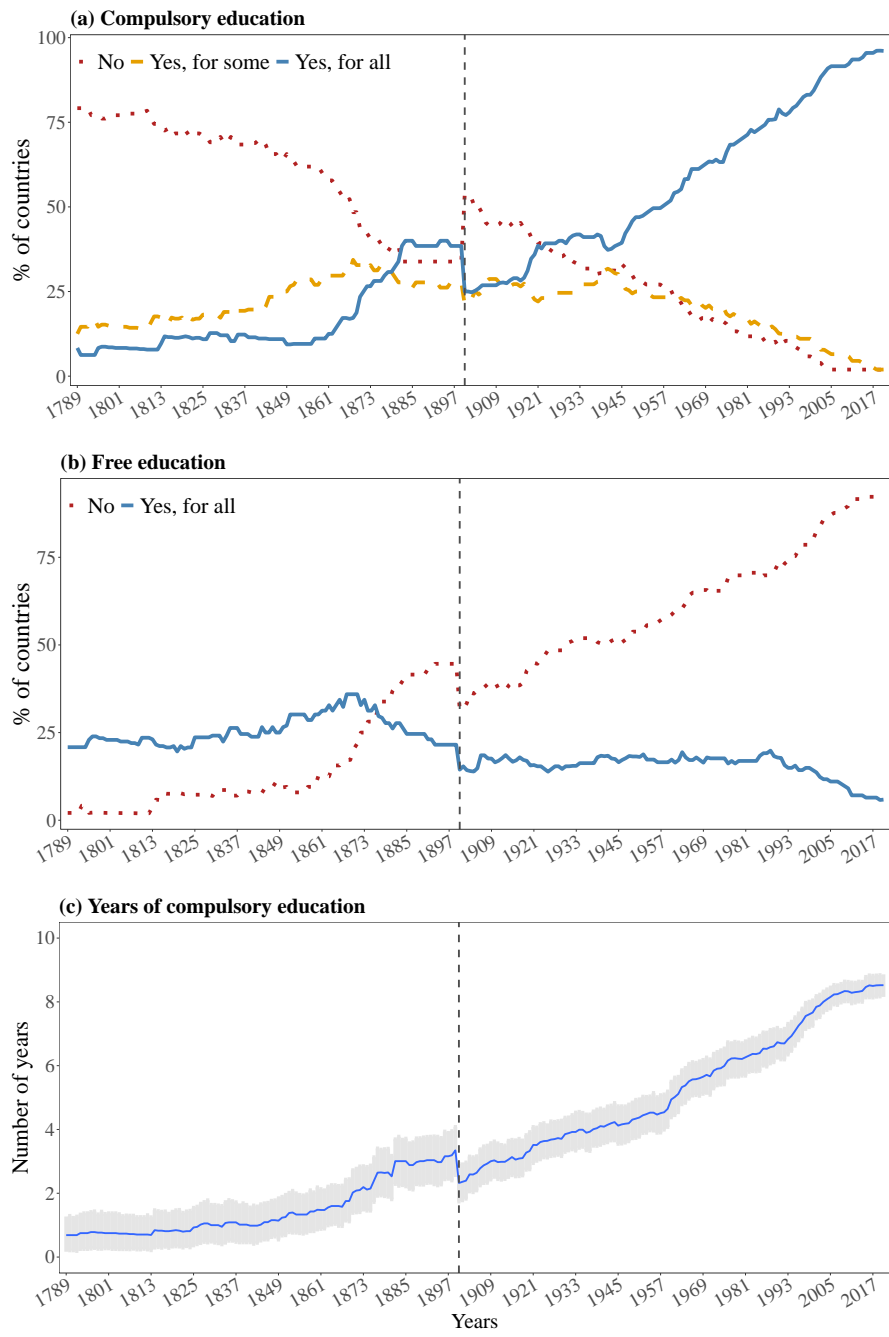


Figure 3. Compulsory education trends

Let us turn now to education content, and more specifically the presence of mandated standalone civics education courses or ideological training at different levels.¹⁰ Figure 4 (panel a) shows that most countries had mandated standalone civic courses during the late 19th century, either at the primary, secondary, or tertiary level. This share remained quite stable before increasing fast after WWII. As of 2020, 60% of countries mandated standalone civics education courses or ideological training in primary *and* secondary education.

Regarding the courses' ideological content, our data reveal rich cross-country and over-time variation. Figure 4 (panel b) collapses the original eight-category variable into three: (1) civics courses without any clear ideological profile, (2) courses or training centered on democratic institutions and norms (including human rights), and (3) courses with another clear ideological profile (i.e., nationalist ideology, regime or leader-specific ideology, or an ideology centered on religion or ethnicity). Before the Cold War ended, the latter type dominated, whereas courses focusing on democratic norms were much less frequent (and about equal in numbers to courses without any clear profile). After the Cold War, however, there has been a rapid increase in the share of countries with civics courses centered on democratic norms – at present, this characterizes the majority of countries, including several electoral authoritarian countries. Thus, democracy not only became a more prominent regime type after the 1980s, but so did the teaching of its core values and functioning in civics classes around the world.

The EPSM data also show a gradual but significant expansion of *public* education since the early 19th century. Figure 5 (panel a) shows that national or sub-national governments have increasingly funded and managed primary and secondary schools. Pri-

¹⁰We define civics education as the teaching of political topics that relate to the functioning or values of the current political regime or a school of thought that has the status of an “official” ideology (e.g., mentioned in the constitution).

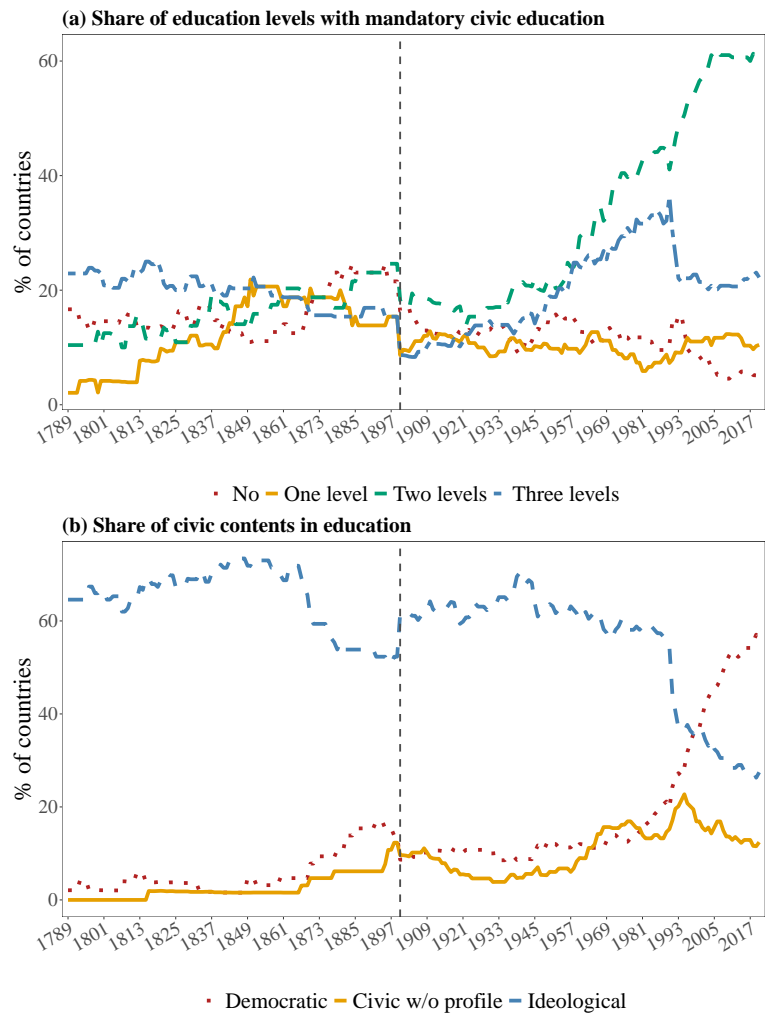


Figure 4. Ideological training trends.

Note: Ideological civics training here refers to civics courses with a profile pertaining to nationalist ideology, regime or leader-specific ideology, or an ideology centered on religion or ethnicity, but does not include courses with a democratic/human rights profile or courses without a clear ideological profile.

vate actors, especially the church, played a key role in operating schools before 1848. After observing a reduced role in many states during the 19th century, private actors were once again prominent in the early 20th century. Yet, the relatively high frequency of privately funded and operated systems in this era relates to the expansion of the sample, with several African and Asian colonies being included after 1900. In many colonies, education services were chiefly provided by religious actors or a mix of public and private actors.

Together with the expansion of publicly funded and managed education systems, we can observe the expansion of state-determined curricula from Figure 5. Before the mid-19th century, states typically did not intervene in the design and implementation of school curricula. From 1860-1945, we observe substantial cross-country heterogeneity in terms of who determined curricula. After 1945, state-determined curricula but also (to a somewhat lesser extent) partly state-determined and partly sub-national government-determined curricula (“shared curricula”) became the dominant practices. And national governments collaborating (or at least sharing responsibilities) with sub-national state actors in determining curricula became even more frequent after the Soviet Union fell, although having state-determined curricula remains the modal practice still today.

Next, Figure 6 (panel a) indicates that, after the 1880s, the most common provider of teacher training is the state. Yet, throughout history (and especially before 1930), many countries have had their teachers trained by non-state actors such as religious groups, international organizations, or foreign countries. Currently, this pertains to almost one-fourth of countries. Panel b shows trends in teacher training, focusing on the requirements for becoming a teacher and the source of teacher training. Concerning

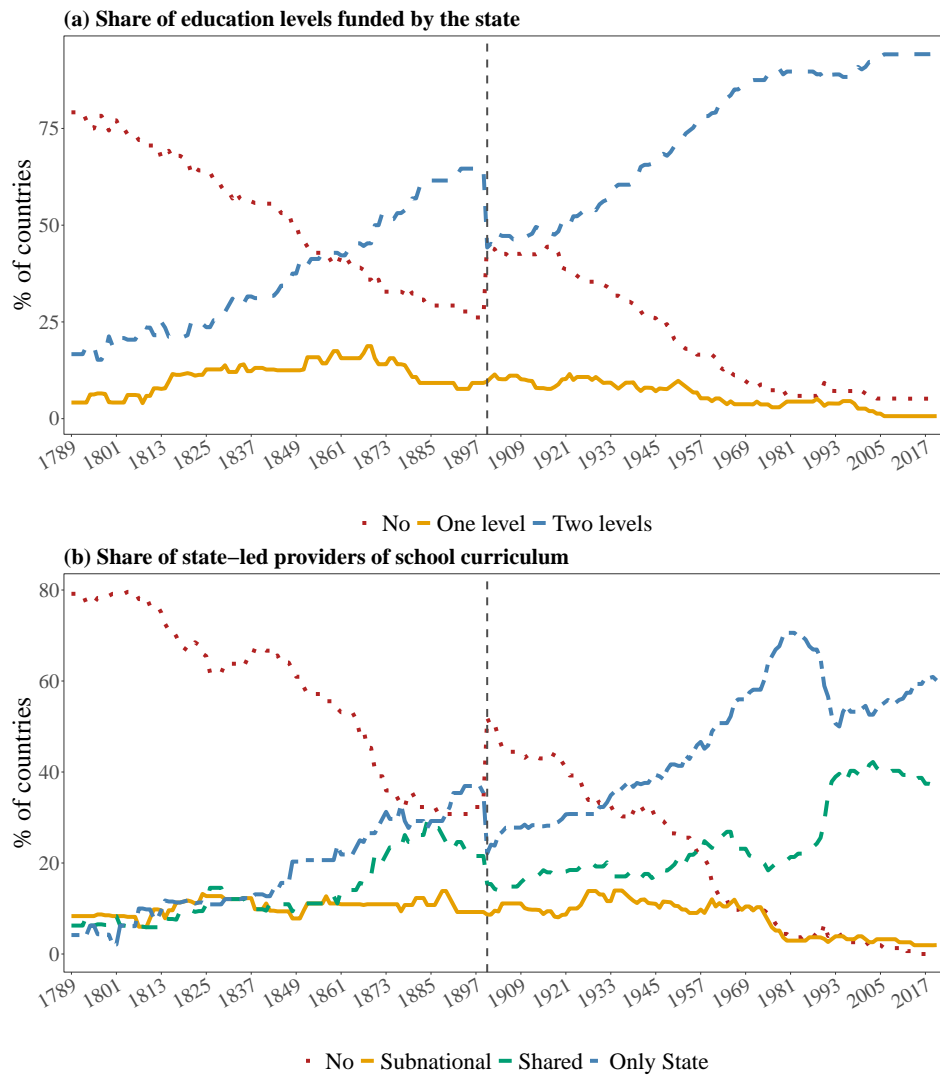


Figure 5. School autonomy.

Note: Panel a focuses on compulsory education and counts primary and secondary education as the two levels of education.

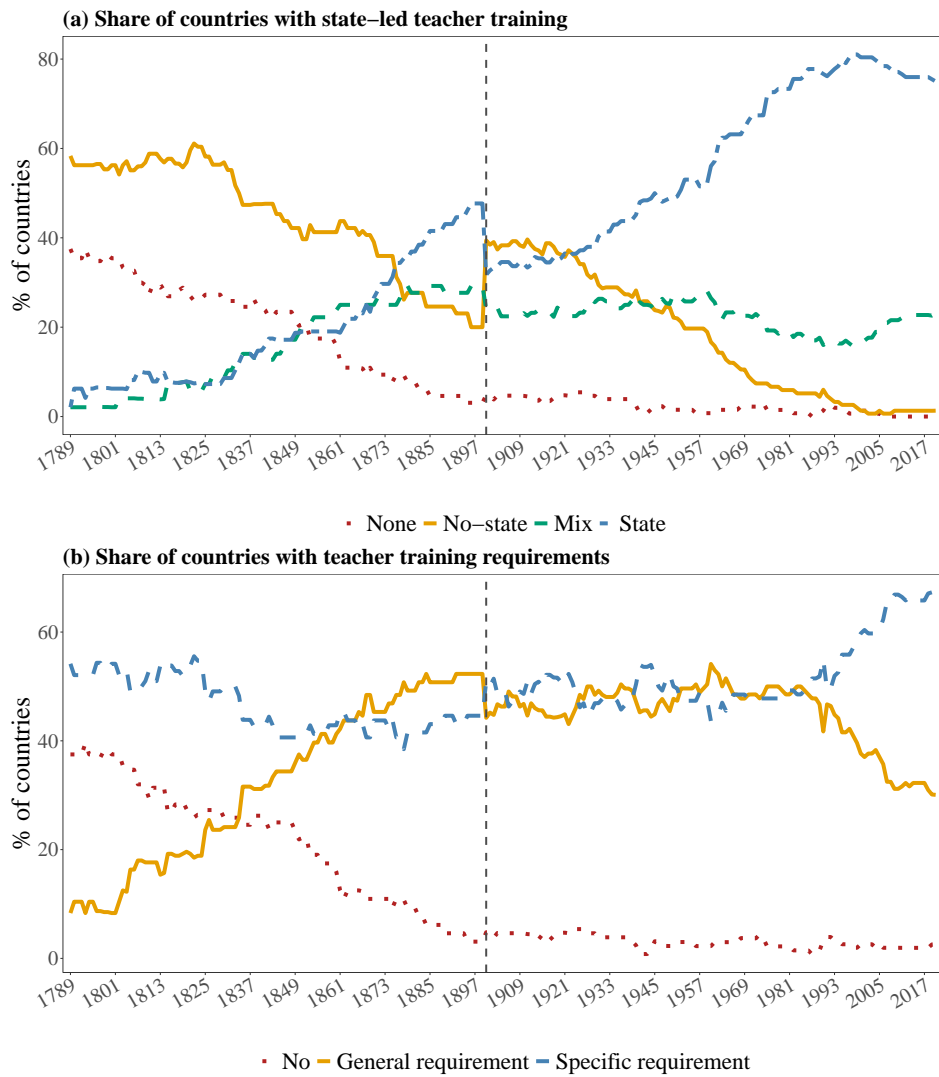


Figure 6. Teacher training.

Note: By “mix” we mean training provided by both private and state actors. “Specific requirements” refer, for example, to a particular degree or type of education, such as a master degree in pedagogical training “General requirements” means that teachers must hold some university degree.

requirements, up until around 1850, most countries either did not require any degree to take on the teacher role, or they had very specific requirements.

Regarding the latter category, most teachers with specific requirements were priests or had to have theological training to become a teacher. This was the case, for example, in the United Kingdom. In the following years, and until the 1960s, our data display an about equal split between countries with specific requirements (a master degree in pedagogical training issued by the state became increasingly common) and general requirements (e.g., teachers must hold some university degree). Since the 1960s, more countries have demanded specific requirements for candidate teachers, and fewer countries have had (only) general requirements.

Applications

In this section, we use EPSM data to replicate some of the core findings of Paglayan (2021) on a larger and geographically more diverse sample of countries before presenting some interesting nuances and extensions using different EPSM variables on specific features of the education system. Studying 33 countries from Europe and Latin America, Paglayan (2021) provides compelling evidence that democratization did not spur the establishment of compulsory education, which typically preceded the first democratization episode. Yet, we do not know whether similar patterns exist in countries on other continents, which typically democratized later or have yet to democratize, and where our EPSM data show that compulsory education tended to arrive later than in Europe and Latin America.

Yet, in line with Paglayan’s original finding, Figure 7 shows that, in most cases globally, compulsory education came before the first instance of democratization.¹¹ We note, however, that there are exceptions to the general pattern, mainly located in Sub-Saharan Africa (e.g., Uganda, Namibia, Lesotho, Botswana) or the post-Soviet space (e.g., Estonia, Armenia).

Next, we explore this relationship more systematically by relying on a recently developed Diff-in-Diff estimator for staggered treatments as introduced by Liu, Wang and Xu (2022). More specifically, we rely on the matrix completion estimator developed by Athey et al. (2021) that allows for unit- and time heterogeneous effects.¹² Figure 8 shows these results, which support the null finding by Paglayan (2021); democratization does not significantly increase the likelihood of a country enacting compulsory education. We find fairly similar results when we instead consider years of compulsory education – even though there are some indications of a positive relationship after three to four decades – or use the EPSM coding for compulsory education for specific groups (see Appendix D). Results are also a bit stronger, indicating a positive (but not very robust) effect once substituting our main measure of democracy with the binary democracy measure based on Skaaning et al. (2015)’s Lexical index. In sum, there are some indications of a positive relationship between democratization and compulsory education in some specifications, but we mostly replicate the null findings also found by Paglayan (2021).

¹¹We rely on a cut-off of 0.4 on V-Dem’s (0-1) Polyarchy index to distinguish autocracies from democracies, giving a threshold for counting democratization episodes that is relatively low but not too far from the most widely used dichotomous measures (see Balz, Vassalai and Hicken, 2022; Coppedge and Wilson, 2022; Pemstein and von Römer, 2022). The result from the later Diff-in-Diff models is robust to using an alternative cut-off (0.5) on Polyarchy and comparable when using a binary measure distinguishing regimes with and without competitive elections, based on the Lexical Index of Electoral Democracy from Skaaning, Gerring and Bartusevičius (2015) in Appendix D.

¹²Relying on the interactive fixed effects estimator produces similar results.



Figure 7. Timing of compulsory education and first democratization episode. *Note:* Red dots indicate dates of first democratization (V-Dem’s Polyarchy index ≥ 0.4). Crosses reflect the date of compulsory education for everyone and triangles compulsory education for certain groups. Countries are grouped by continent.

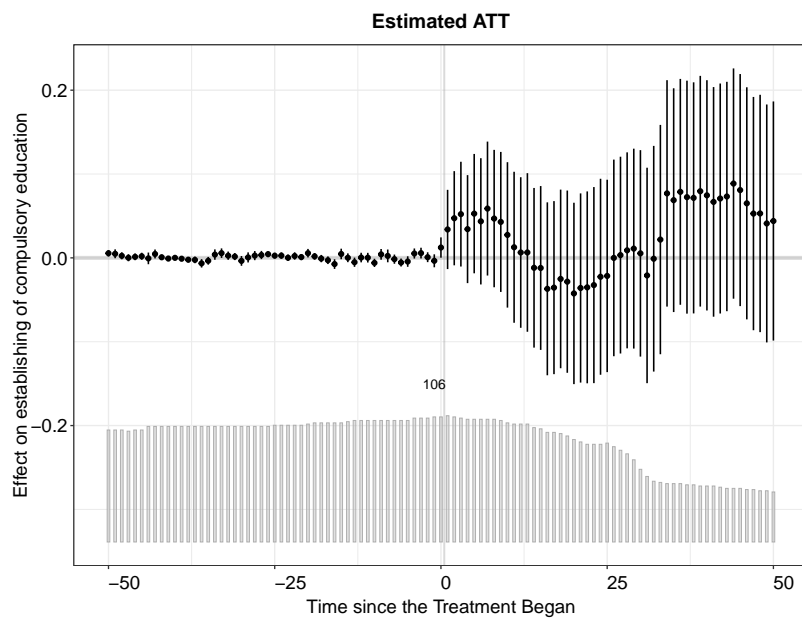


Figure 8. Results from staggered Diff-in-Diff estimation following Liu, Wang and Xu (2022). Outcome: establishment of compulsory education; treatment: first democratization as defined above. *Note:* Time to treatment is limited to ± 50 years.

However, the lacking relationship between the timing of the first democratization episode and compulsory education does not imply that democratization is irrelevant to how education systems evolve, more generally, or to what education policies are implemented. Neither do the results imply that democracies and autocracies – which are typically poorer and differ systematically from democracies also in other relevant regards – are similar even when it comes to compulsory education. This is highlighted in Figure 9, which displays yearly mean scores surrounded by 95% confidence intervals for each regime category for four EPSM variables.

We divide countries into autocracies and democracies, relying again on the 0.4 cut-off on V-Dem’s (0-1) Polyarchy index.¹³ One important observation from Figure 9 is that the resemblance between the “typical” autocratic- and democratic education systems depends on what feature of the system we consider and the period under consideration. For the scope of civic education (panel b), measured by the number of education levels with mandated civics courses, democracies and autocracies resemble each other from the end of the 19th century to the end of World War II (WWII). Autocracies surpassed democracies in the scope of civics training during the Cold War, and afterward, the gap narrowed. We also observe similarities concerning centralized state control over education at the primary and secondary levels (panel d).¹⁴ On average, democracies had relatively more centralized control early on in the 20th century, but the regime averages have later converged so that there is virtually no difference after the Cold War.

¹³In Appendix E.2, we replicate the results using a 0.5-score threshold. Results are also fairly similar when using a binary measure distinguishing regimes with and without competitive elections, based on the Lexical Index of Electoral Democracy from Skaaning, Gerring and Bartusevičius (2015).

¹⁴The centralized education system measure is an index, summing across binary indicators identifying whether the state funds primary and secondary education (`operate_prim` & `operate_sec`) and the establishment of centralized curricula (`edu_power`). The index is normalized and ranges from 0-1 so that a 1-score means that the state imposes school curricula and manages the funding of schools. See Appendix B for details.

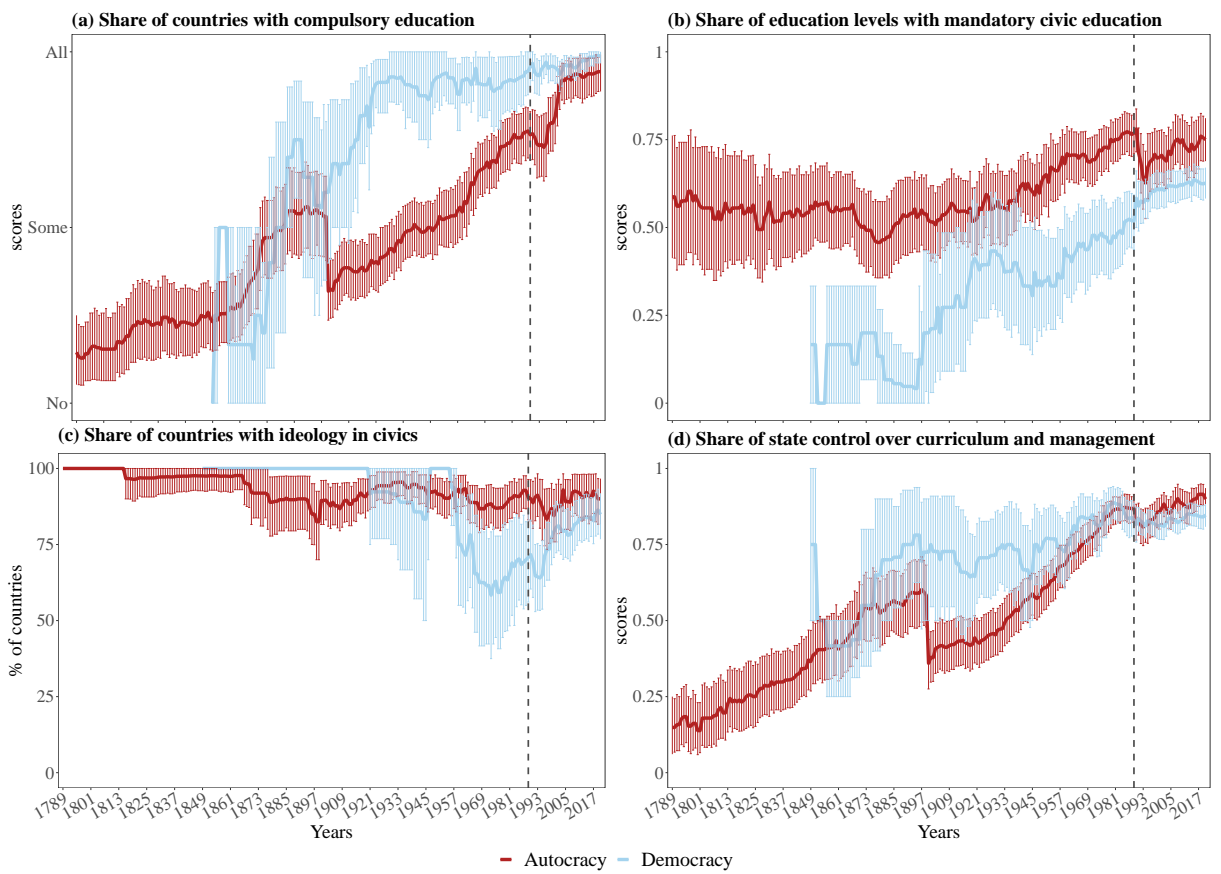


Figure 9. Education system developments by regime type.

Note: "Ideological civics courses" here include democracy-oriented courses. In Appendix E.1, we present plots distinguishing between courses with different ideological content.

For other measures, there are clear regime differences. Almost all democracies provide compulsory education for everyone, at least after the mid-1920s (panel a). In contrast, autocracies have gradually increased the scope of compulsory education throughout the century, and only after 1989 have 70% autocracies provided compulsory education for everyone. Another clear difference emerges for the ideological content of civics courses (panel b). Historically, almost all autocracies (> 85%) have consistently had ideological content in mandated civics courses in primary and/or secondary education, while this number fluctuated over time for democracies. Specifically, democracies and autocracies were, on average, very different after WWII, but have converged again in recent decades as more democracies have implemented mandated civics courses, typically with democracy and human rights as ideological content.

Conclusion

We have introduced our new dataset, Education Policies and Systems across Modern History (EPSM). EPSM incorporates 21 variables on compulsory education, the ideological content of education, political control of education institutions, as well as the training of teachers. EPSM covers 156 countries and some time series extend from 1789 to the present. This unprecedented empirical coverage will hopefully enable researchers to arrive at better-founded descriptions of the historical development of education systems and policies globally. We also believe that this dataset can be used to address a range of important questions pertaining to education and its roles in shaping (and being shaped by) various social, economic, and political developments. In particular, the many variables concerning the ideological content and political control over education systems, from primary school to university, means the dataset is especially useful for

addressing, e.g., questions of indoctrination and how different political regimes may use education systems to achieve different goals.

In this paper, we discussed how the EPSM data, which are mainly measuring de jure features of education policies and systems, were collected before discussing and assessing its reliability and validity characteristics. We also discussed several descriptive patterns, for instance, the almost two centuries-long and gradual rise to global dominance of the free, compulsory education model and the sharp post-Cold War increase in mandated civics courses espousing democratic norms. We also used the data to replicate and extend analyses on the (lack of any clear) relationship between countries experiencing democratization episodes and the introduction of compulsory education.

Numerous other important questions can be (re-)addressed in future research by using these new data. To take one example, the exact nature and strength of the relationship between education and economic growth remain unsettled empirical questions, despite the centrality of human capital accumulation for economic growth in several theoretical growth models (see, e.g., Lucas, 1988; Mankiw, Romer and Weil, 1992; Romer, 1990). Previous studies have highlighted the importance of education measurement for this relationship, for instance suggesting that proxies tapping into the quality of education and what kids actually learn in school are far better predictors of growth than education quantity measures, such as average years of schooling (Hanushek and Woessmann, 2008, 2012). Using data on the ideological content of education or the system for training teachers, panel analyses using our data could help shed further light on what aspects of education policies matter for growth and what types of education are (and are not) productivity-enhancing.

Another example pertains to the historical legacies of colonialism. Several studies have proposed, and to varying extents documented empirically, how political institu-

tions (e.g., Acemoglu, Johnson and Robinson, 2001) or even social policies (Grünewald, 2021) adopted during European colonization persisted and have had longer-term ramifications for the nature of institutions and policies in many Latin American, Asian, and African countries. Similarly, the type of education systems and policies adopted during the colonial era might have had lasting effects, and existing studies have indicated that this is certainly the case for colonial-era education of elites in Sub-Saharan African countries (Ricart-Huguet, 2021). Our data could allow researchers to assess hypotheses on historical persistence due to colonial legacies in a wider set of countries and assess the extent of persistence or change for different aspects of education systems.

The above-mentioned examples are still only a small subset of the topics that may be addressed by using our new data. Hence, we hope and believe that EPSM will allow scholars with different research interests and from different disciplines to arrive at better-founded empirical answers to old but unsettled questions pertaining to education. In addition, we hope that EPSM can contribute to spurring new and more specific questions about the historical development, causes, and effects of education policies and systems.

References

- Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review* 91(5):1369–1401.
- Ainley, John and Ralph Carstens. 2018. “Teaching and learning international survey (TALIS) 2018 conceptual framework.”
- Angrist, Noam, Harry Anthony Patrinos and Martin Schlotter. 2013. “An expansion of a global data set on educational quality : a focus on achievement in developing countries.” *Policy Research working paper* WPS 6536.
- Ansell, Ben and Johannes Lindvall. 2013. “The Political Origins of Primary Education Systems: Ideology, Institutions, and Interdenominational Conflict in an Era of Nation-Building.” *American Political Science Review* 107(3):505–522.
- Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens and Khashayar Khosravi. 2021. “Matrix completion methods for causal panel data models.” *Journal of the American Statistical Association* 116(536):1716–1730.
- Balz, Samuel, Fabricio Vasselai and Allen Hicken. 2022. “An unexpected consensus among diverse ways to measure democracy.” *Democratization* 29(5):814–837.
- Barro, Robert J. and Jong Wha Lee. 2013. “A new data set of educational attainment in the world, 1950–2010.” *Journal of Development Economics* 104:184–198.
- Bromley, Patricia, Rie Kijima, Lisa Overbey, Jared Furuta, Minju Choi and Heitor Santos. 2023. “World Education Reform Database, vi.”

Coppedge, Michael, John Gerring Carl Henrik Knutsen Staffan I. Lindberg Jan Teorell David Altman Michael Bernhard Agnes Cornell M. Steven Fish Lisa Gastaldi Haakon Gjerløw Adam Glynn Sandra Grahn Allen Hicken Katrin Kinzelbach-Kyle L. Marquardt Kelly Mc-Mann Valeriya Mechkova Pamela Paxton Daniel Pemstein Johannes von Römer Brigitte Seim Rachel Sigman Svend-Erik Skaaning Jeffrey Staton Eitan Tzelgov Luca Uberti Yi-ting Wang Tore Wig and Daniel Ziblatt. 2022a. *V-Dem Codebook v12*. Varieties of Democracy (V-Dem) Project.

Coppedge, Michael, John Gerring Carl Henrik Knutsen Staffan I. Lindberg Jan Teorell Kyle L. Marquardt Juraj Medzihorsky Daniel Pemstein Nazifa Alizada Lisa Gastaldi Garry Hindle Josefina Pernes Johannes von Römer Eitan Tzelgov Yi-ting Wang and Steven Wilson. 2022. *V-Dem Methodology v12*. Varieties of Democracy (V-Dem) Project.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell and Lisa Gastaldi. 2022. *V-Dem Country Coding Units, v.12*. Varieties of Democracy (V-Dem) Project.

Coppedge, Michael, John Gerring Carl Henrik Knutsen Staffan I. Lindberg Jan Teorell Nazifa Alizada David Altman Michael Bernhard Agnes Cornell M. Steven Fish Lisa Gastaldi Haakon Gjerløw Adam Glynn Sandra Grahn Allen Hicken-Garry Hindle Nina Ilchenko Katrin Kinzelbach Joshua Krusell Kyle L. Marquardt Kelly Mc-Mann Valeriya Mechkova Juraj Medzihorsky Pamela Paxton Daniel Pemstein Josefina Pernes Oskar Rydén Johannes von Römer Brigitte Seim Rachel Sigman Svend-Erik Skaaning Jeffrey Staton Aksel Sundström Eitan Tzelgov Yi-ting Wang Tore Wig Steven Wilson and Daniel Ziblatt. 2022b. *V-Dem [Country-Year/Country-Date] Dataset, v.12*. Varieties of Democracy (V-Dem) Project.

- Dahlum, Sirianne. 2017. *Schooling for dissent? Education, autocratic regime instability and transitions to democracy* PhD thesis Department of Political Science, University of Oslo.
- Dahlum, Sirianne. 2019. "Students in the Streets: Education and Nonviolent Protest." *Comparative Political Studies* 52(2):277–309.
- Frank-Michael Kuhlemann, *Niedere Schulen. 1991. 1970 – 1918 :von der Reichsgründung bis zum Ende des Ersten Weltkriegs.* München : Beck chapter 4, pp. 221–3.
- Grünwald, Aline. 2021. "From Benefits and Beneficiaries: The Historical Origins of Old-Age Pensions From a Political Regime Perspective." *Comparative Political Studies* 54(8):1424–1458.
- Hanushek, Eric A and Ludger Woessmann. 2008. "Education and economic growth." *Chapters in Economics* .
- Hanushek, Eric A and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth* 17(4):267–321.
- Lee, Jong-Wha and Hanol Lee. 2016. "Human capital in the long run." *Journal of Development Economics* 122:147–169.
- Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1):69–105.
- Liu, Licheng, Ye Wang and Yiqing Xu. 2022. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* .

- Lucas, Robert E. 1988. "On the Mechanisms of Economic Development." *Journal of Monetary Economics* 22:3–42.
- Mankiw, Gregory N., David Romer and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *The Quarterly Journal of Economics* 107(2):407–437.
- Murtin, Fabrice and Romain Wacziarg. 2014. "The democratic transition." *Journal of Economic Growth* 19:141–181.
- Neundorf, Anja, Eugenia Nazrullaeva, Ksenia Northmore-Ball, Katerina Tertytchnaya and Wooseok Kim. 2023. "Varieties of Indoctrination (V-Indoc): Introducing a Global Dataset on the Politicization of Education and the Media."
- Paglayan, Agustina S. 2021. "The Non-Democratic Roots of Mass Education: Evidence from 200 Years." *American Political Science Review* 115(1):179–198.
- Paglayan, Agustina S. 2022. "Education or Indoctrination? The Violent Origins of Public School Systems in an Era of State-Building." *American Political Science Review* p. 1–16.
- Pemstein, Daniel, Kyle L. Marquardt Eitan Tzelgov Yi-ting Wang Juraj Medzihorsky Joshua Krusell Farhad Miri and Johannes von Römer. 2022. *The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data*. Working Paper No. 21. 7th edition. University of Gothenburg: Varieties of Democracy Institute.
- Petterson, Lars. 1991. "1842, 1822 or 1882? What is it that should be celebrated? Research on education." *Journal for analysis and debate* 18(4):22–27.

- Ricart-Huguet, Joan. 2021. "Colonial Education, Political Elites, and Regional Political Inequality in Africa." *Comparative Political Studies* 54(14):2546–2580.
- Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98(5):71–102.
- Sen, Amartya. 1999. *Development as Freedom*. New York: Anchor Books.
- Skaaning, Svend-Erik, John Gerring and Henrikas Bartusevičius. 2015. "A Lexical Index of Electoral Democracy." *Comparative Political Studies* 48(12):1491–1525.
- Theis, Lisa. 2021. *Bildungssystem im Deutschen Kaiserreich. Bedeutung des Militärs für Schule und Unterricht*. GRIN Verlag.
- Thyne, Clayton. 2006. "ABC's, 123's, and the Golden Rule: The Pacifying Effect of Education on Civil War, 1980–1999." *International Studies Quarterly* 50(4):733–754.
- Westberg, Johannes. 2017. *Funding the Rise of Mass Schooling: The Social, Economic and Cultural History of School Finance in Sweden, 1840–1900*. Cham: Palgrave Macmillan.

Online Appendices for “Education Policies and Systems across Modern History: A Global Dataset”

These online appendices contain additional descriptive analyses (Online Appendix A), such as time series lines for the compulsory education history of each country in the EPSM dataset (A₁) or other indicators (A₂). In Appendix B, we detail the Centralization index, which we built from several EPSM variables and used for descriptive analysis in the paper. Appendix C illustrates several of the EPSM measures by plotting and discussing the historical development of the education systems in two countries, Spain and Russia. In Appendix D, we replicate the diff-in-diff analysis on the first democratization episodes and potential effects on education systems, but on slightly different compulsory education outcome variables than the “compulsory education for all” measure used in the main paper. Appendix E replicates some of the analyses, using an alternative democracy measure than the one used in the paper. Appendix F is the EPSM codebook and Appendix G is the Rules-of-thumb-for-coding document.

A Additional descriptions analyses

A.I Compulsory education across and within countries

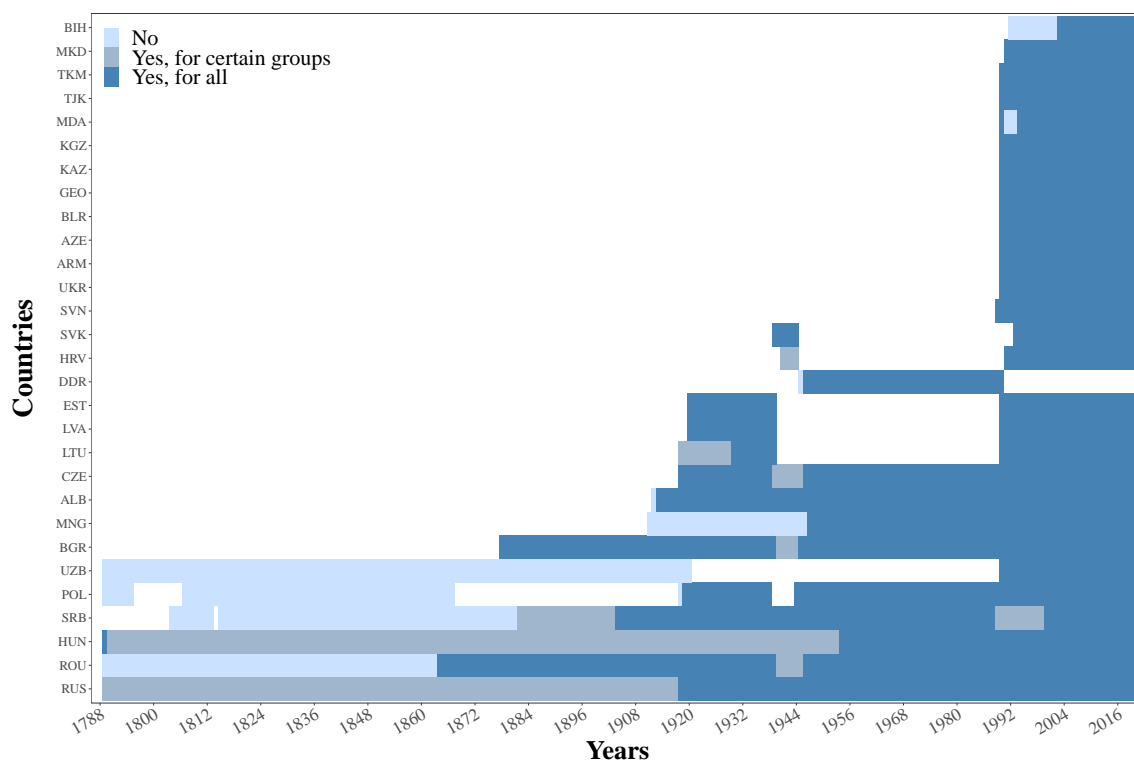


Figure SM.I. Eastern Europe and Central Asia

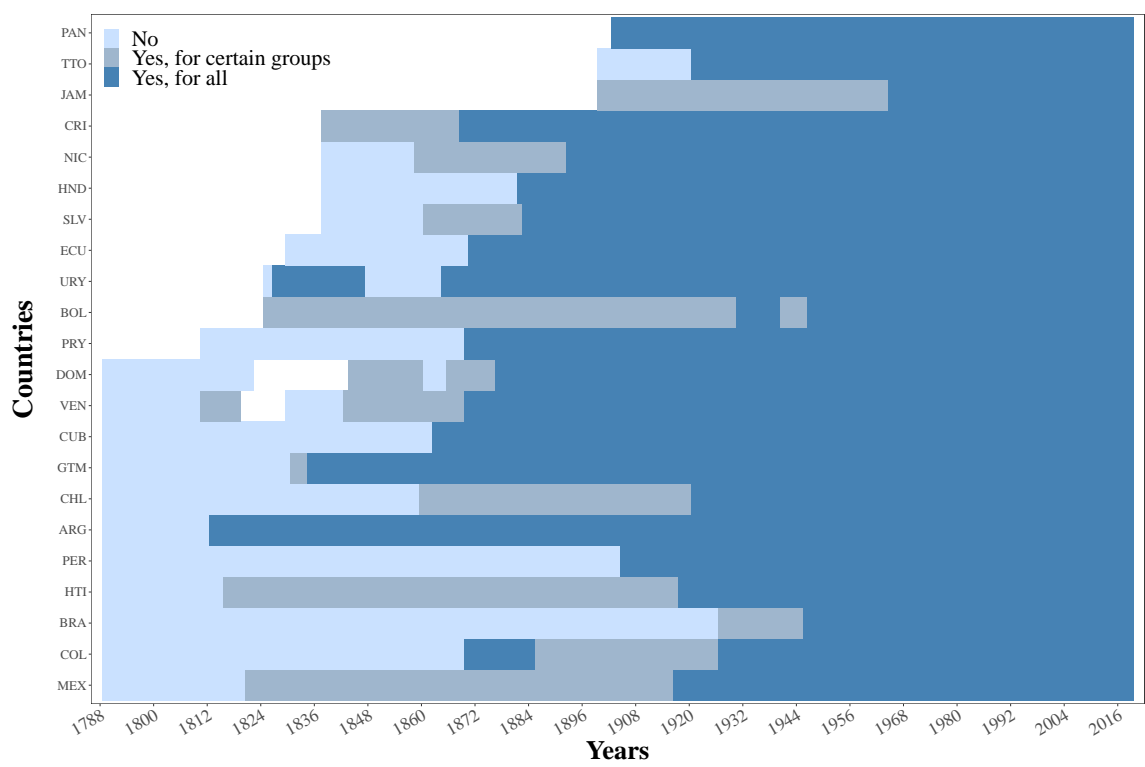


Figure SM.2. Latin America and the Caribbean

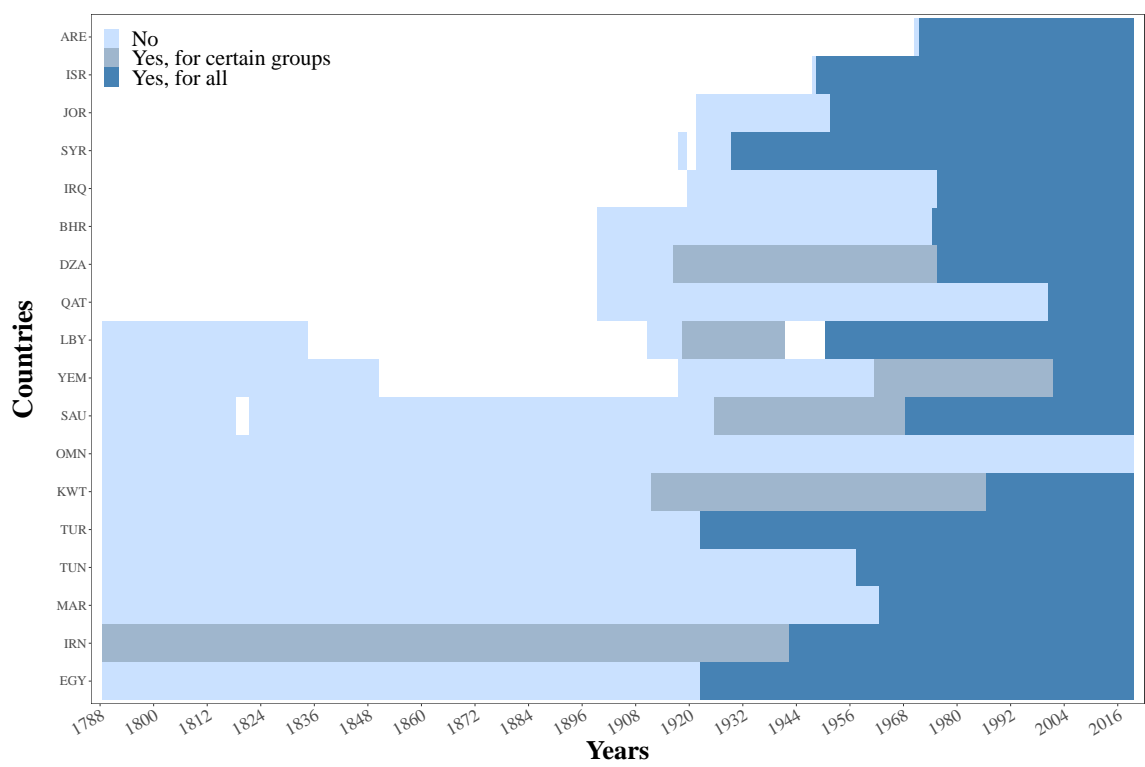


Figure SM.3. The Middle East and North Africa

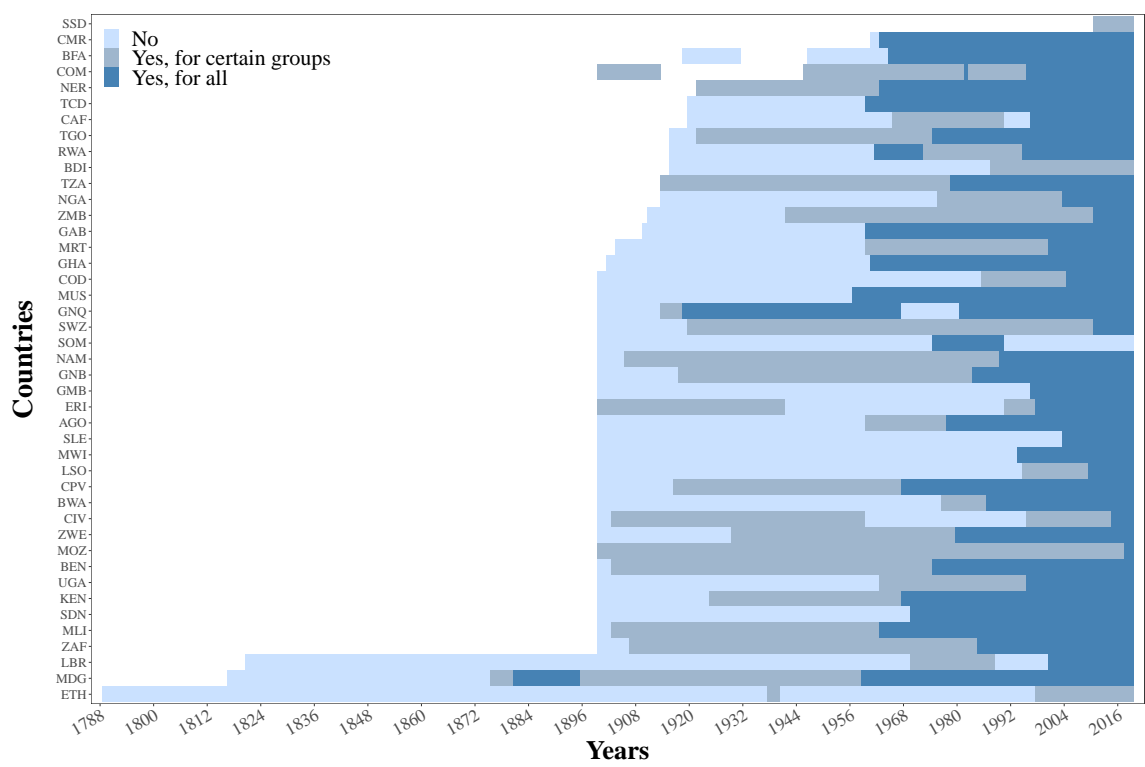


Figure SM.4. Sub-Saharan Africa

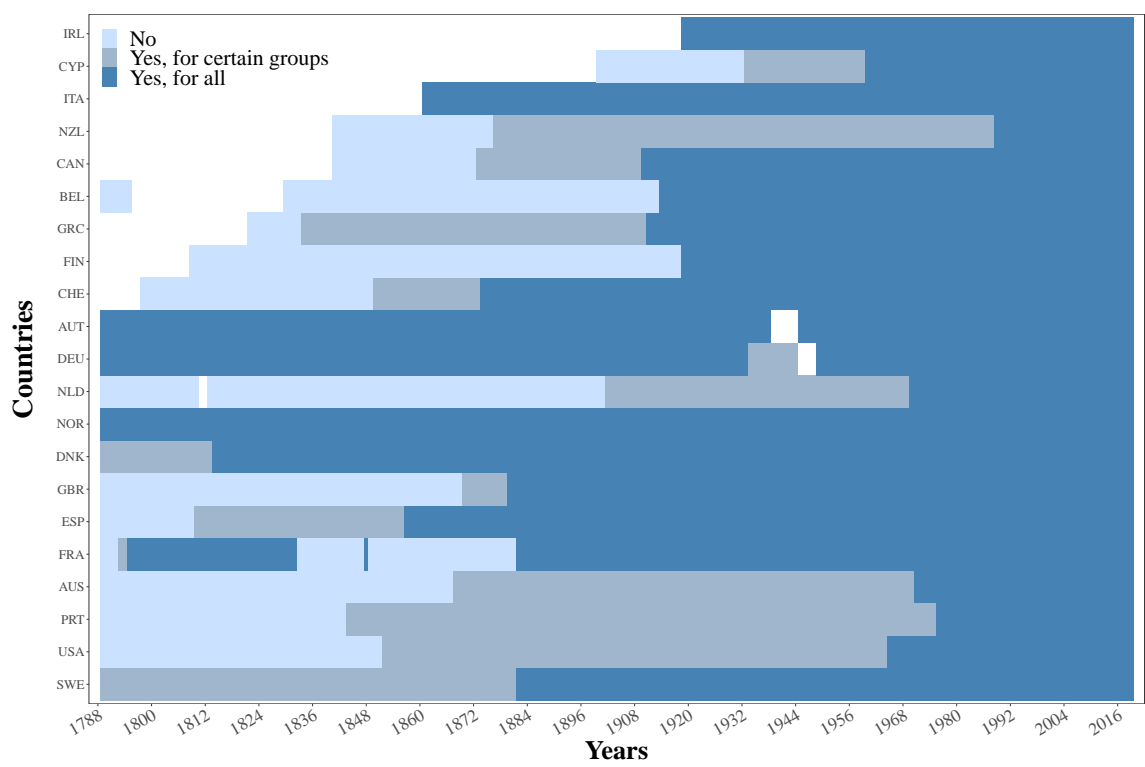


Figure SM.5. Europe and North America

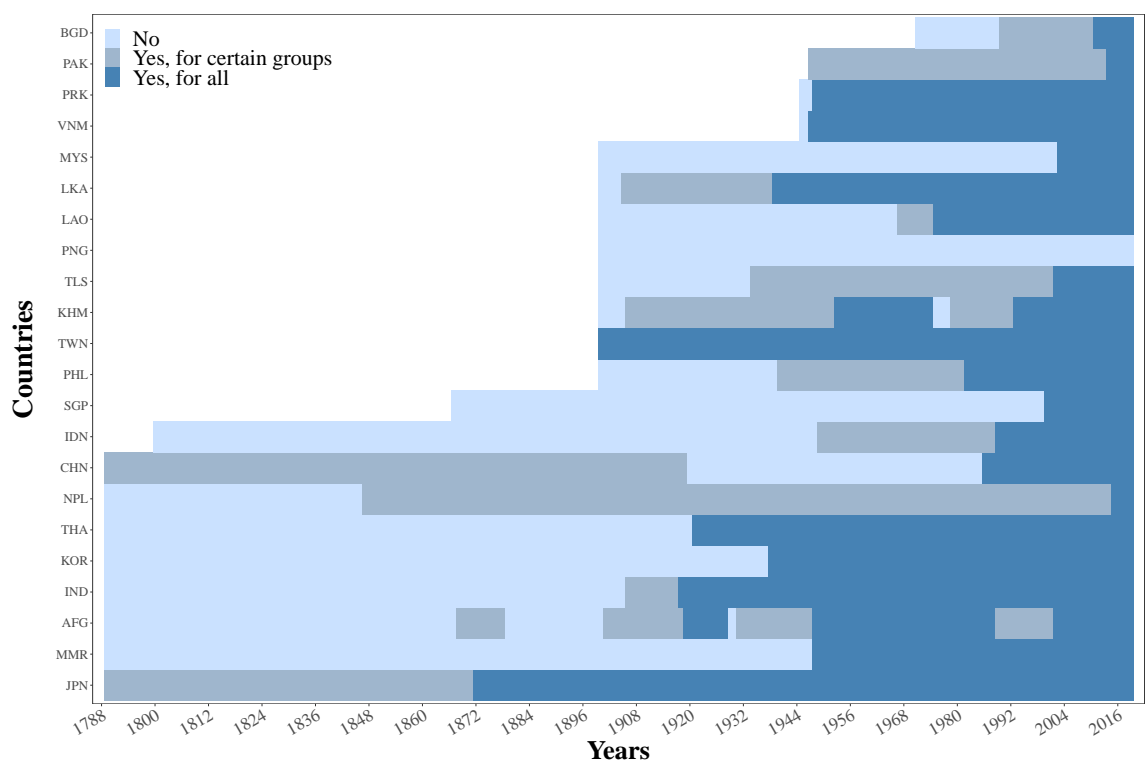


Figure SM.6. Asia and Pacific

A.2 Patterns of education worldwide

This subsection contains plots of additional patterns of education systems and policies worldwide. Figure SM.7 plots types of civic education courses, focusing on the most frequent categories: civic education without a clear ideological profile as well as civic education centered on, respectively, democracy, religion, nationalism, and regime-specific ideologies (e.g., communism, fascism, among others). One very prolific pattern is the gradual decline of religion in making citizens over the course of modern history. In contrast, courses centered on nationalist values have been on the rise, especially after the 19 century.

Historically, neither civics courses centered on democratic values or courses centered on more regime-specific ideologies were especially prevalent. Indeed, only around 10% of countries had democracy-centered courses before World War I, and the number was even lower for regime-specific ideology courses. From the interwar period, however, we observed a gradual increase in regime-specific education until reaching its peak in 1990, when 30% of states included this type of civic training. After the fall of the Soviet Union, there was a sharp decrease in this type of education. This shift went together with an even more marked increase in civics training centered on democratic values.

Next, Figures SM.8-13 show variation across and within countries regarding bans on education subjects or topics. In most regions of the world, such kind of censorship is not really the norm despite notable exception (e.g., in Europe, Nazi Germany and its censorship body in education or the post-second world war German education system, where elements violating the constitution, such as Nazi ideology or militarism in the education, were banned). Yet, overall, we found few explicit bans on education as manifested in education laws, decrees, and constitutions worldwide.

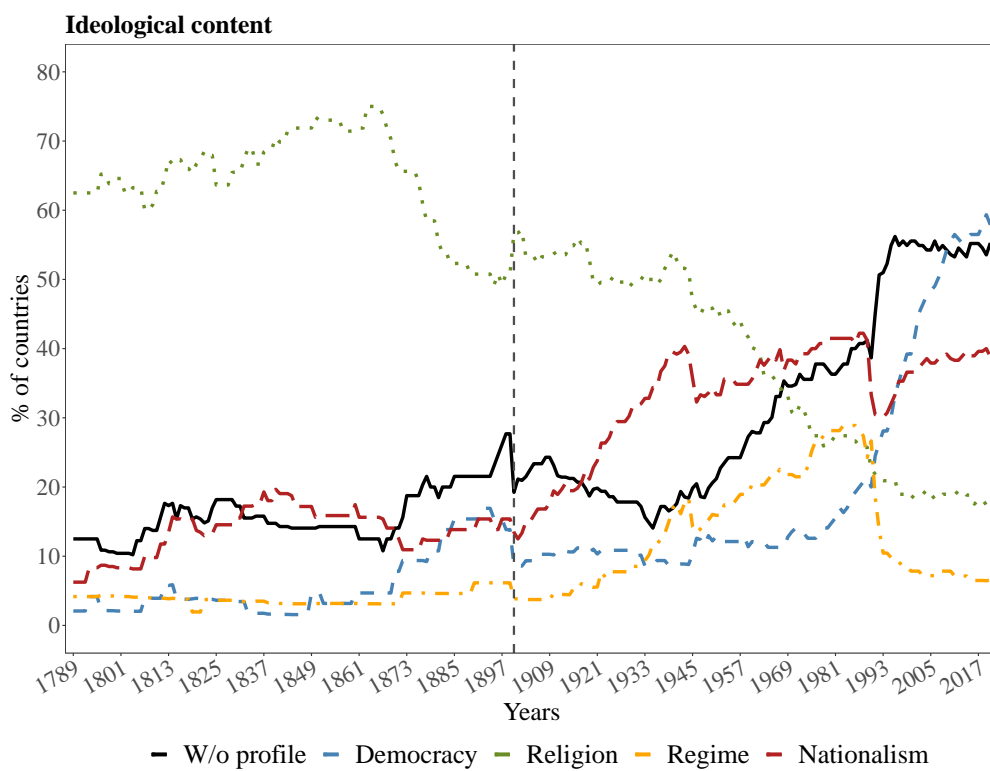


Figure SM.7. Patterns of civics profile

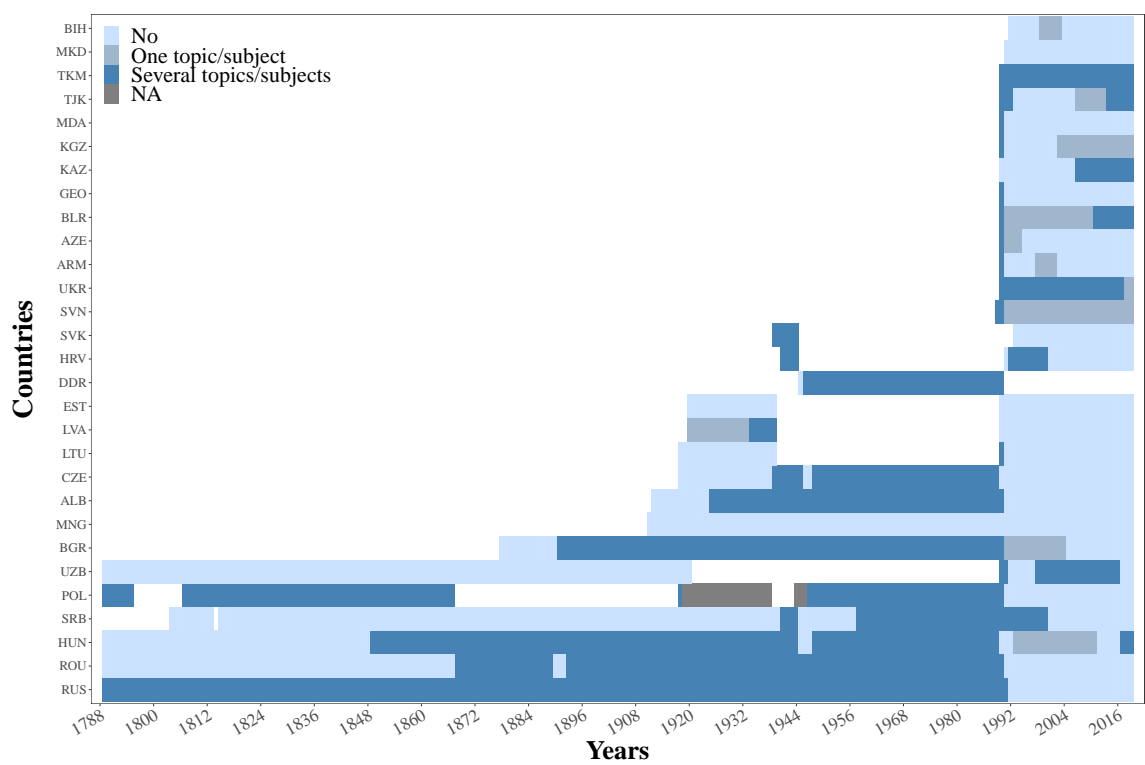


Figure SM.8. Subject bans: Eastern Europe and Central Asia

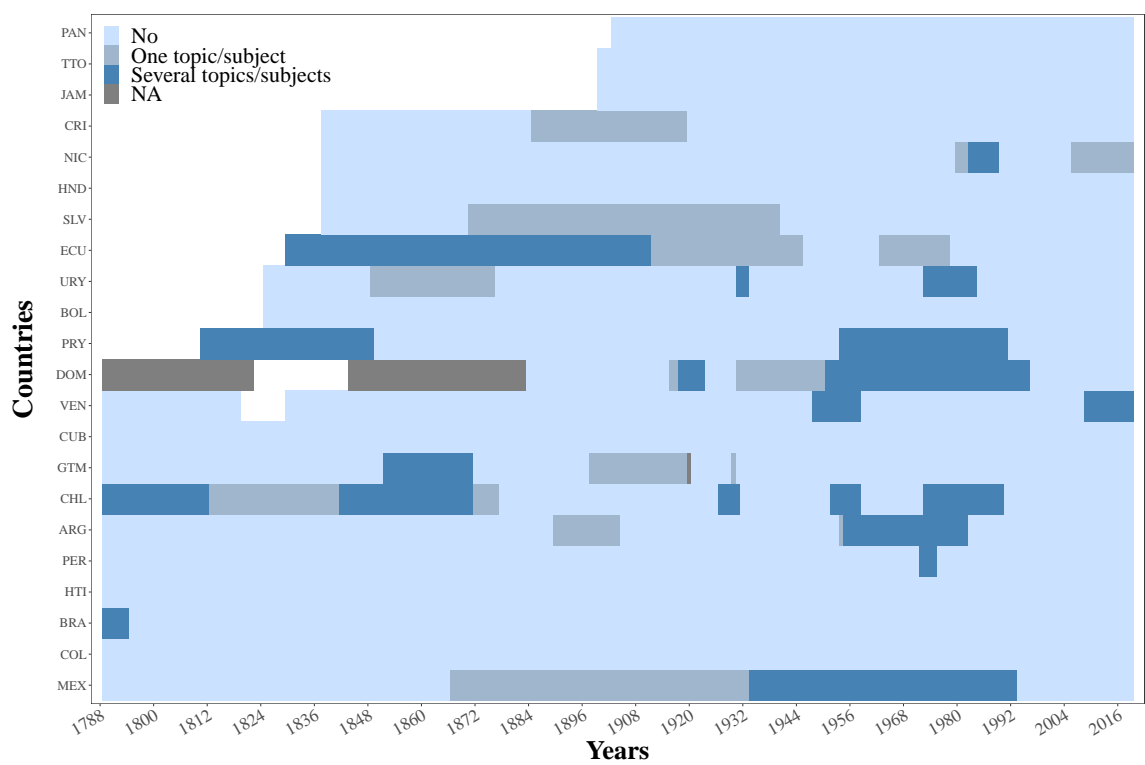


Figure SM.9. Subject bans: Latin America and the Caribbean

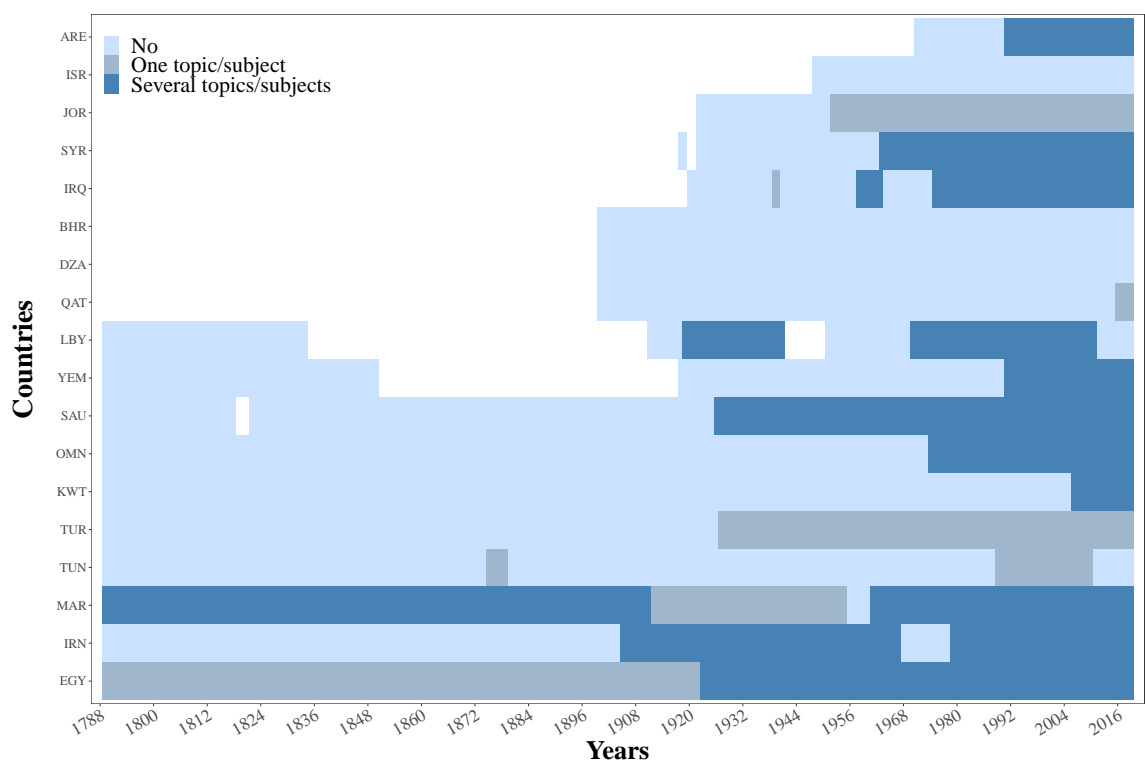


Figure SM.10. Subject bans: The Middle East and North Africa

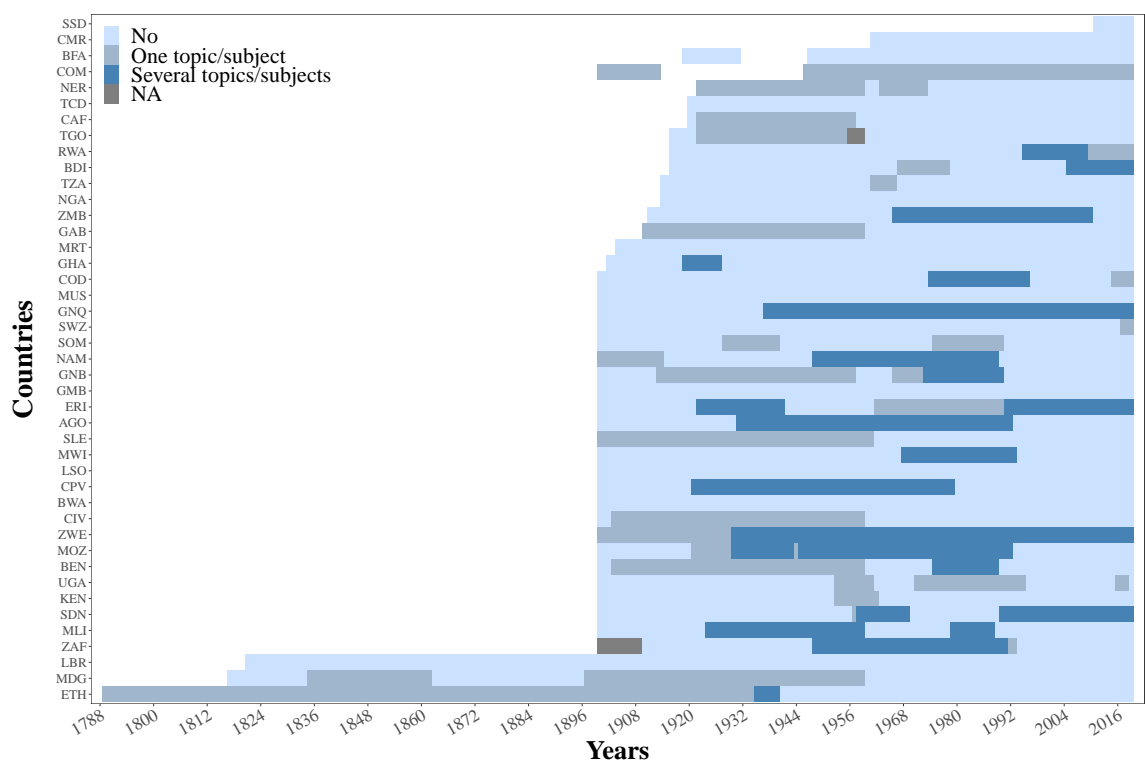


Figure SM.II. Subject bans: Sub-Saharan Africa

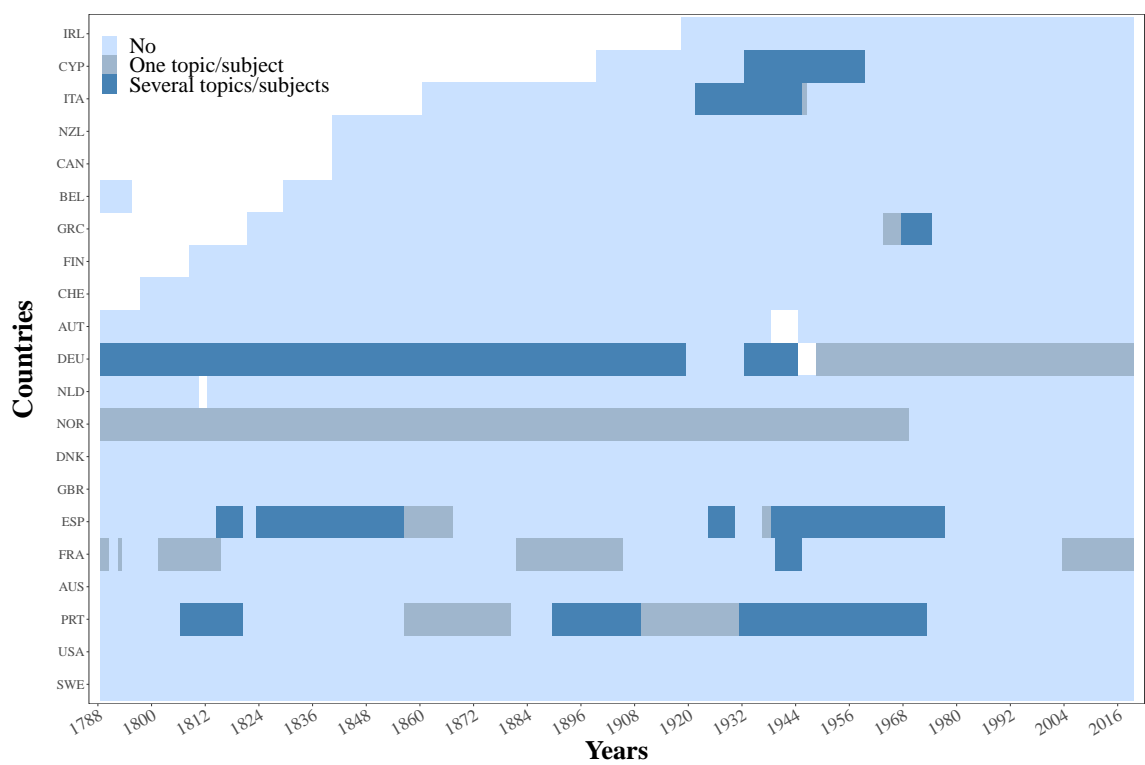


Figure SM.12. Subject bans: Europe and North America

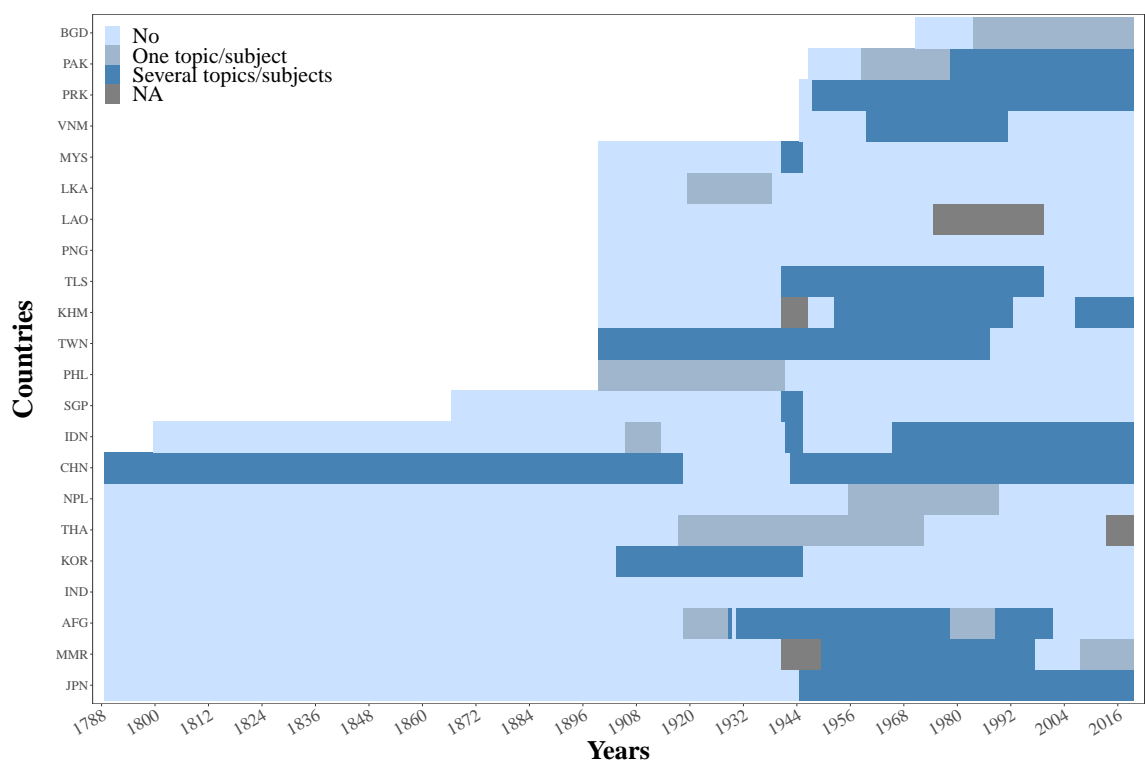


Figure SM.13. Subject bans: Asia and Pacific

B Centralization measure

The education system centralization measure used in the main paper is an index, summing across binary indicators identifying whether the state funds primary and secondary education (`operate_prim` and `operate_sec`) and the establishment of centralized curricula (`edu_power`). The index is normalized and ranges from 0 to 1. The specific operationalization is as follows:

`operate_prim` = 1 (for primary education) and `operate_sec` = 1 (for secondary education) when there is evidence that “the state” funds schools, and 0 otherwise. The “state funds schools” condition occurs when funding comes from: (1) Local government, (2) regional government, (3) national government, (4) private actors and at least one from alternatives 1-3, and (5) two or more of alternatives 1-3. If the state thus funds schools at both the primary and secondary level, the combined score from these two indicators is 2, whereas it is 1 if state funding only applies to one level.

The final indicator, which ranges from 0-2, is a re-coded version of the original EPSM variable `edu_power`, so that: 0 = There is no centralized curricula provided by the national government or by regional government 1 = There is a centralized curriculum provided by a regional government only 2. There is a centralized curriculum provided by a national government only or in collaboration with a sub-national government.

After summing `operate_prim`, `operate_sec` and `edu_power`, we normalized the resulting 0-4 scaled index, so that 0 express the minimum levels of state intervention and 1 the maximum levels of state control over the curricula and management for the final, normalized index.

C Illustrative examples: Spain and Russia

Figure SM.14 plots historical developments of the education systems in two example countries, namely Spain and Russia. More specifically, the figure plots temporal patterns along four key variables contained in EPSM: compulsory education (top-left panel), years of compulsory education (top-right panel), centralized curriculum (bottom-left panel), and coverage of public education (bottom-right panel).

The upper panel in Figure SM.15 plots whether the government bans several subjects (value 3), one or few subjects (value 2), or none (value 1). The bottom panel shows how many levels of civics courses are mandated, where the value 4 denotes that civic courses are mandatory from primary schools to the university, while the value 1 denotes that there are no mandatory civics courses.

Figure SM.15 depicts the share of countries with civics courses of particular kinds, focusing on five types of contents: civics without an ideological profile, democratic norms, religion, regime-specific ideology (e.g., fascism), and nationalism. These are the most frequent categories in the EPSM ideology scheme. In Figure SM.16, we consider the same civic course ideology measures but zoom in on Spain and Russia. As the figures show, civic training could have multiple contents. Sometimes we observe courses without an ideological profile and with an ideological profile. This emphasizes the fact that some school curricula include compulsory subjects like sociology or social studies together with courses with a more clear ideological profile, such as "citizenship education," "me and the world," or "Patriotic education."

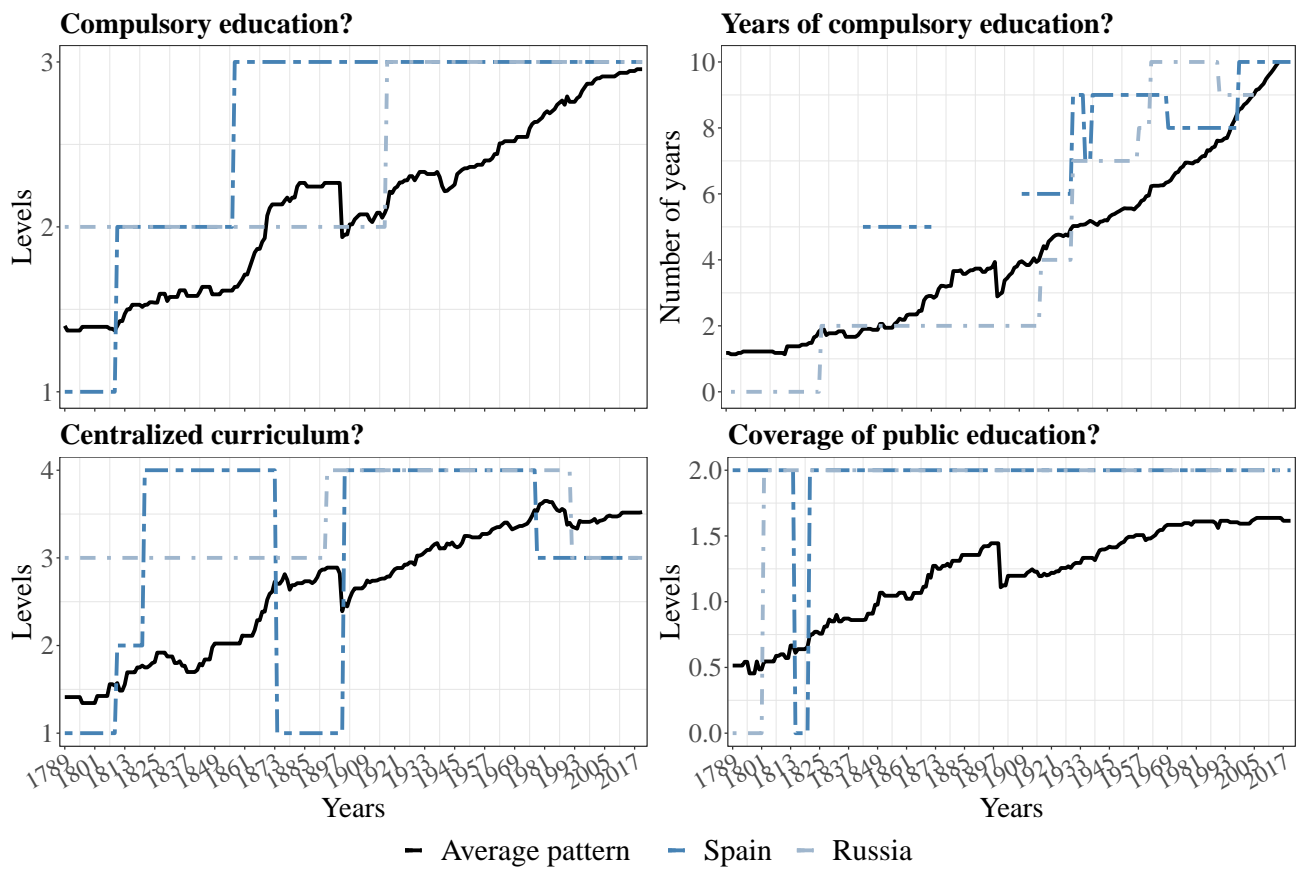


Figure SM.14. Patterns of education systems (a)

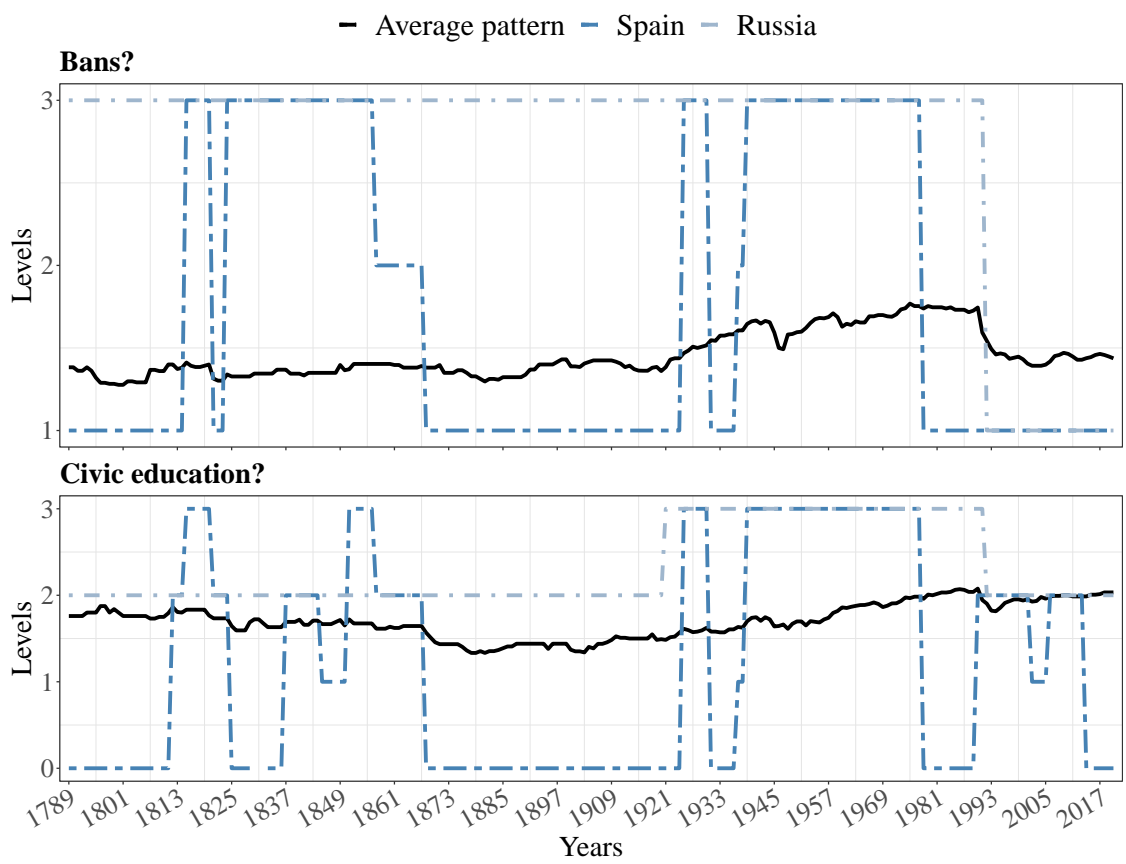


Figure SM.15. Patterns of education systems (b)

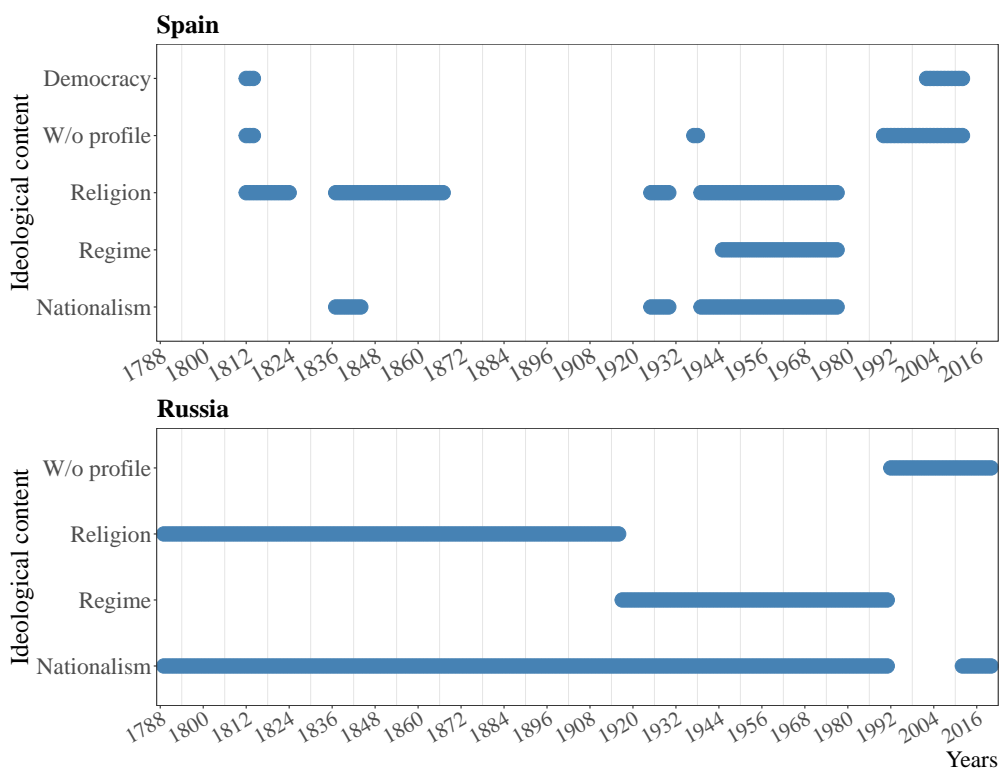


Figure SM.16. Civics profile in Spain and Russia

D Additional Diff-in-Diff models on different compulsory education outcomes

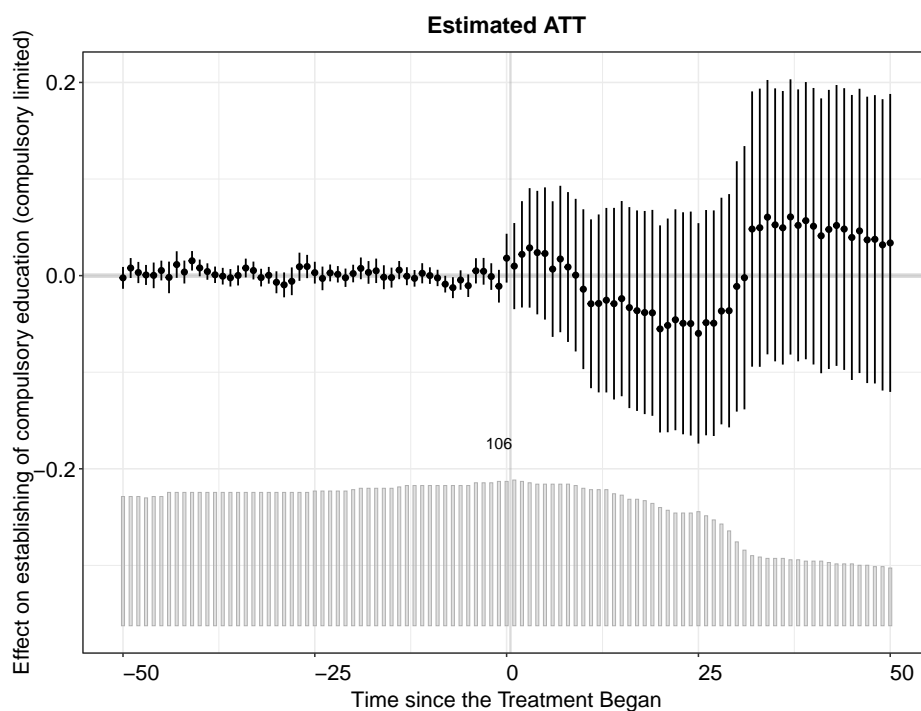


Figure SM.17. Results of a staggered Diff-in-Diff estimation following Liu, Wang and Xu (2022). Outcome: establishment of compulsory education for specific groups; treatment: first democratization (as defined in the paper). *Note:* Time to treatment is limited to ± 50 years.

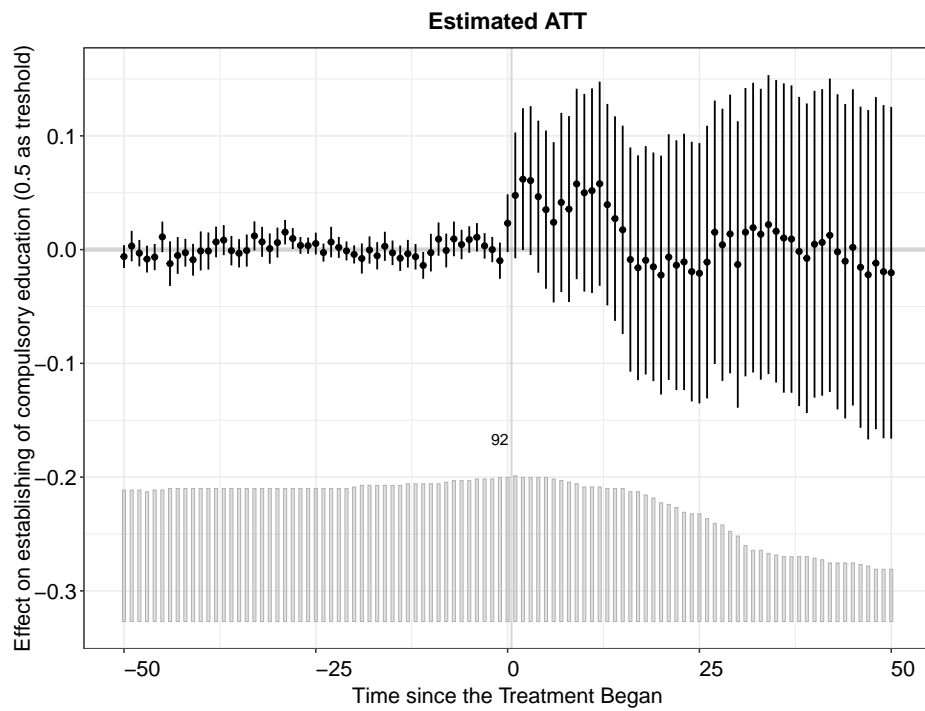


Figure SM.18. Results of a staggered Diff-in-Diff estimation following Liu, Wang and Xu (2022). Outcome: establishment of compulsory education; treatment: first democratization (with having 0.5 as cut-off on V-Dem's (0-1) Polyarchy index). *Note:* Time to treatment is limited to ± 50 years.

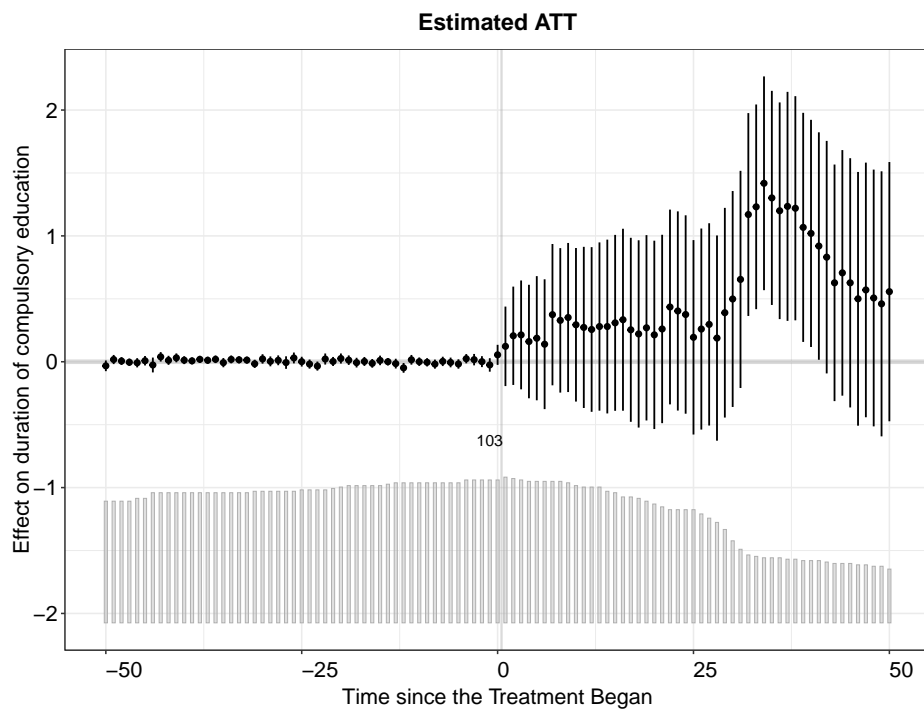


Figure SM.19. Results of a staggered Diff-in-Diff estimation following Liu, Wang and Xu (2022). Outcome: years of compulsory education; treatment: first democratization (as defined in the paper). *Note:* Time to treatment is limited to ± 50 years.

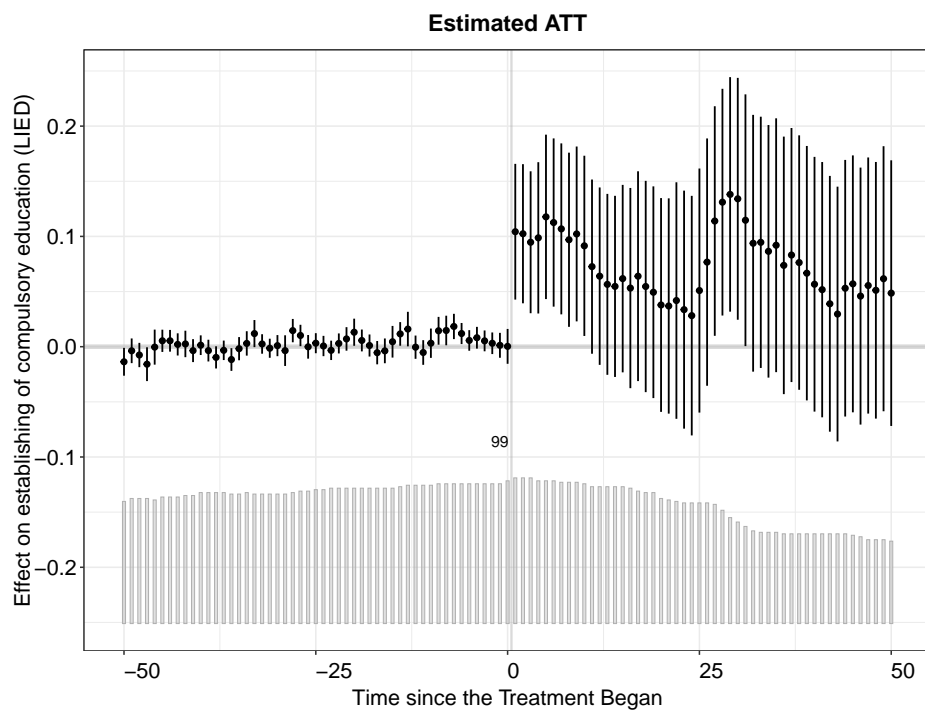


Figure SM.20. Results of a staggered Diff-in-Diff estimation following Liu, Wang and Xu (2022) Outcome: establishment of compulsory education; treatment: first democratization (based on the Lexical Index of Electoral Democracy from Skaaning, Gerring and Bartusevičius (2015)). *Note:* Time to treatment is limited to ± 50 years.

E Patterns of education systems in democracies and autocracies

E.1 Additional descriptive analyses

Figure SM.21 shows fine-grained information about ideology in civic education and subject bans between democracies and autocracies. There are clear differences between both regime types for most of the period covered, especially after the Second World War. One interesting pattern is that most authoritarian regimes include subjects that taught students democratic norms and values post-Cold War. Yet, democratic norms and values are the defining features of education in democracies as civic training in another type of ideologies has been in decline since WWII while civic training without a clear ideological profile is also in decline and within the 20% to 40%. As of 2020, 75% of democracies have standalone civic training in democratic norms and values.

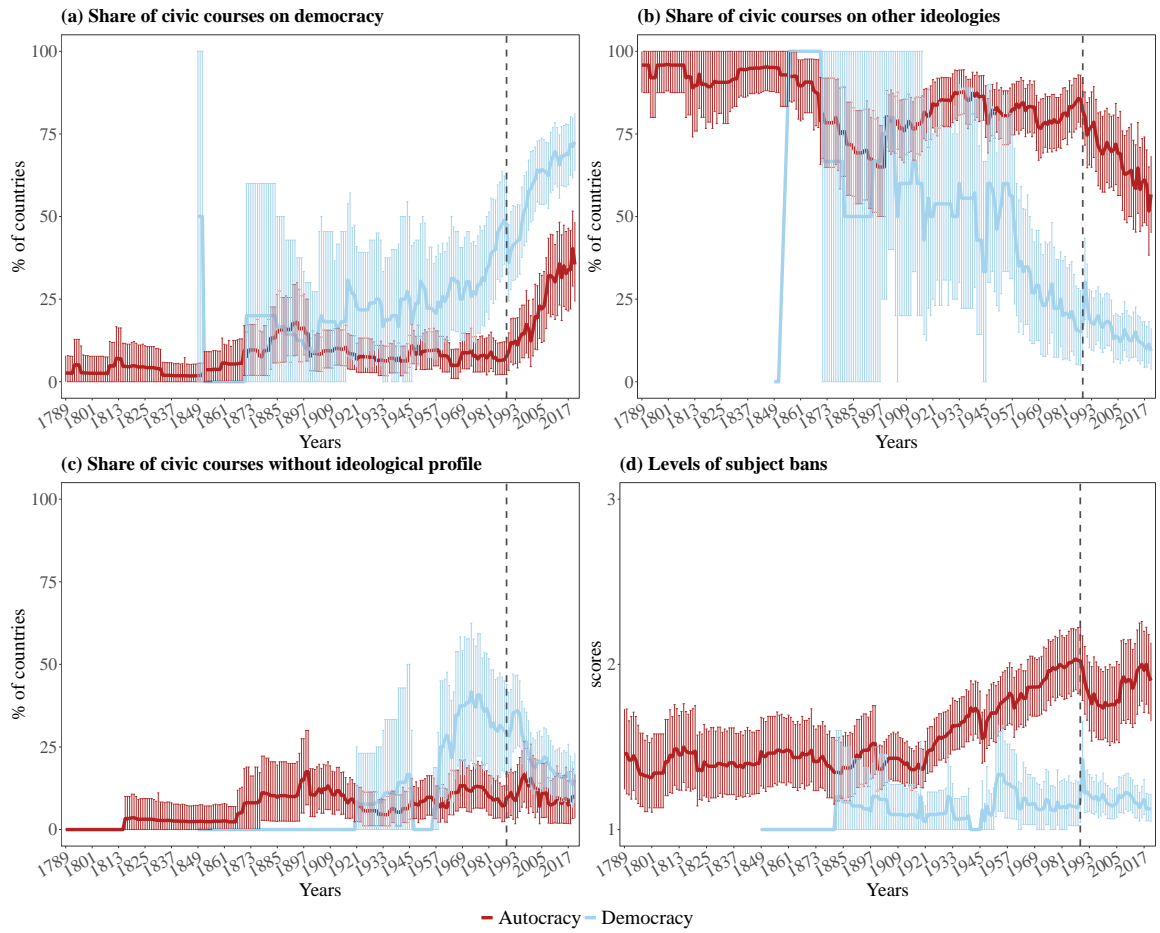


Figure SM.2.1. Patterns of education systems by regime type.

Note: The score 1 on subject bans refers to the absence of explicit laws prohibiting topics or subjects in school. The score 2 denotes that one specific subject or topic is forbidden, while a 3-score denotes that several subjects or topics are forbidden.

E.2 Using an alternative measure to classify democracies and autocracies

Figures SM.22 and SM.23 plot patterns of education system development, separating democracies and autocracies. In the paper, we use a cut-off of 0.4 on V-Dem's Polyarchy index to distinguish regime types, but in these Appendix figures, we use a 0.5 cut-off point on the same index.

We also use Skaaning et al.'s 2015 lexical index of electoral democracy (LIED). More specifically, we define autocracies as those regimes that do not hold competitive elections (below the value of 4 on the original Lexical scale, which ranges from 0-6). As for Polyarchy (and in contrast to other democracy measures), the Lexical index is coded all the way from 1789 for several countries, allowing us to display education system differences for the entire time series also contained in EPSM. However, we are skeptical of using this measure in our application section as we lose 21% of observations. This number of missing values leads to large standard errors when we replicate our staggered Diff-in-Diff estimations.

Despite different measurement procedures, the competitive elections-based measure relying on LIED leads to fairly similar findings as the measure used in the main paper. Specifically, Figure SM.24 shows a substantial overlap between democracies and autocracies regarding the scope of civic education and centralized curricula over time, especially during 1850-1980.

Democracies and autocracies also resemble each other in the type of compulsory education during the same period (upper-left panel). Yet, after the fall of the Soviet Union, we observe that democracies always enforce compulsory education for everyone, while autocracies are slightly less likely to provide such service.

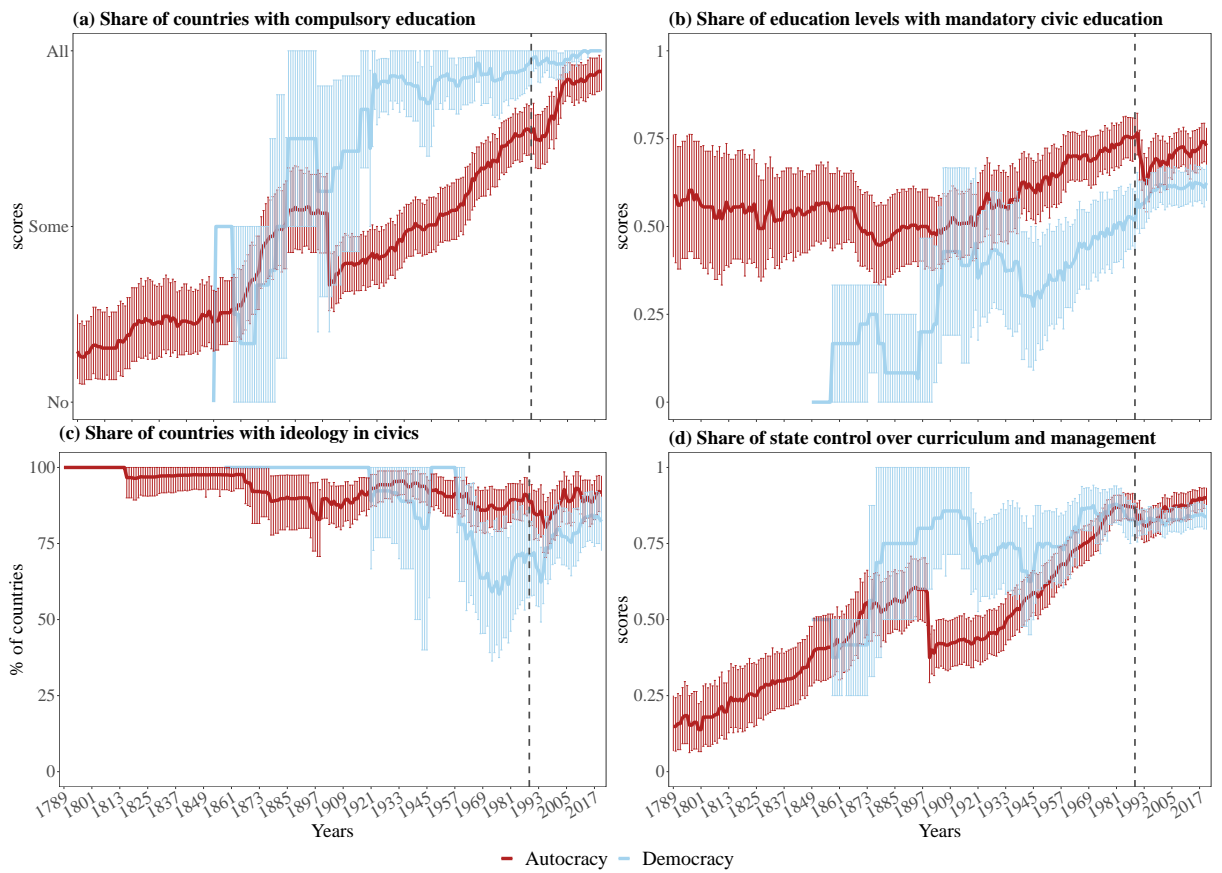


Figure SM.22. Patterns of education systems by regime type (using 0.5-threshold on V-Dem's Polyarchy index to construct binary democracy measure)

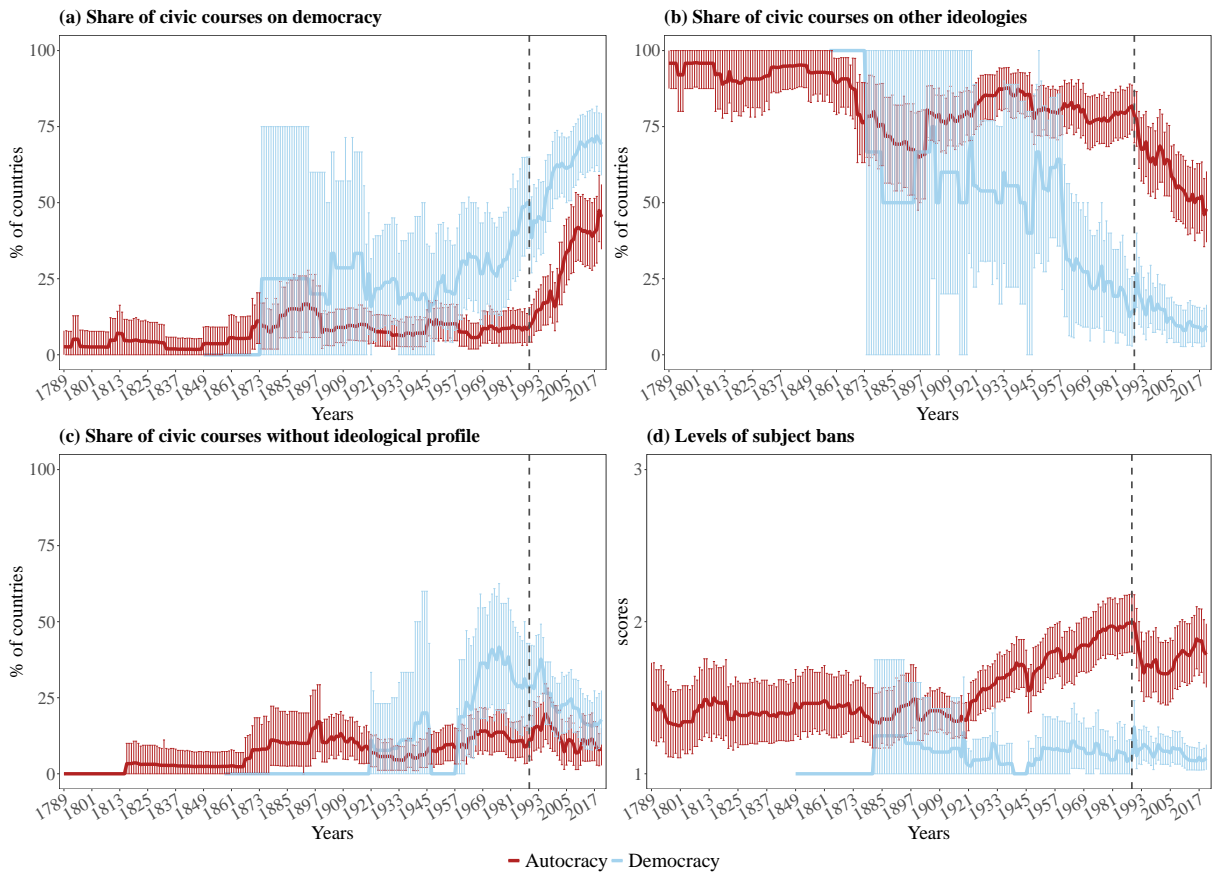


Figure SM.23. Patterns of education systems by regime type (using 0.5-threshold on V-Dem's Polyarchy index to construct binary democracy measure).

Note: The score 1 on subject bans refers to the absence of explicit laws prohibiting topics or subjects in school. The score 2 denotes that one specific subject or topic is forbidden, while a 3-score denotes that several subjects or topics are forbidden.

Regarding our measure of the ideological content in civics classes (lower-left panel), almost all autocracies that had mandated civics classes had ideological content. After WWII, autocracies and democracies differed as fewer democracies included democratic norms and values in their civic training. After 2000, democracies and autocracies behaved similarly pertaining to their mandated civics classes, although some differences exist.

Figure SM.25 on types of civic education and subject bans evolved over time and across regimes show a similar pattern to the one described in the main text. Yet, we observe that differences between regime types are narrowed.

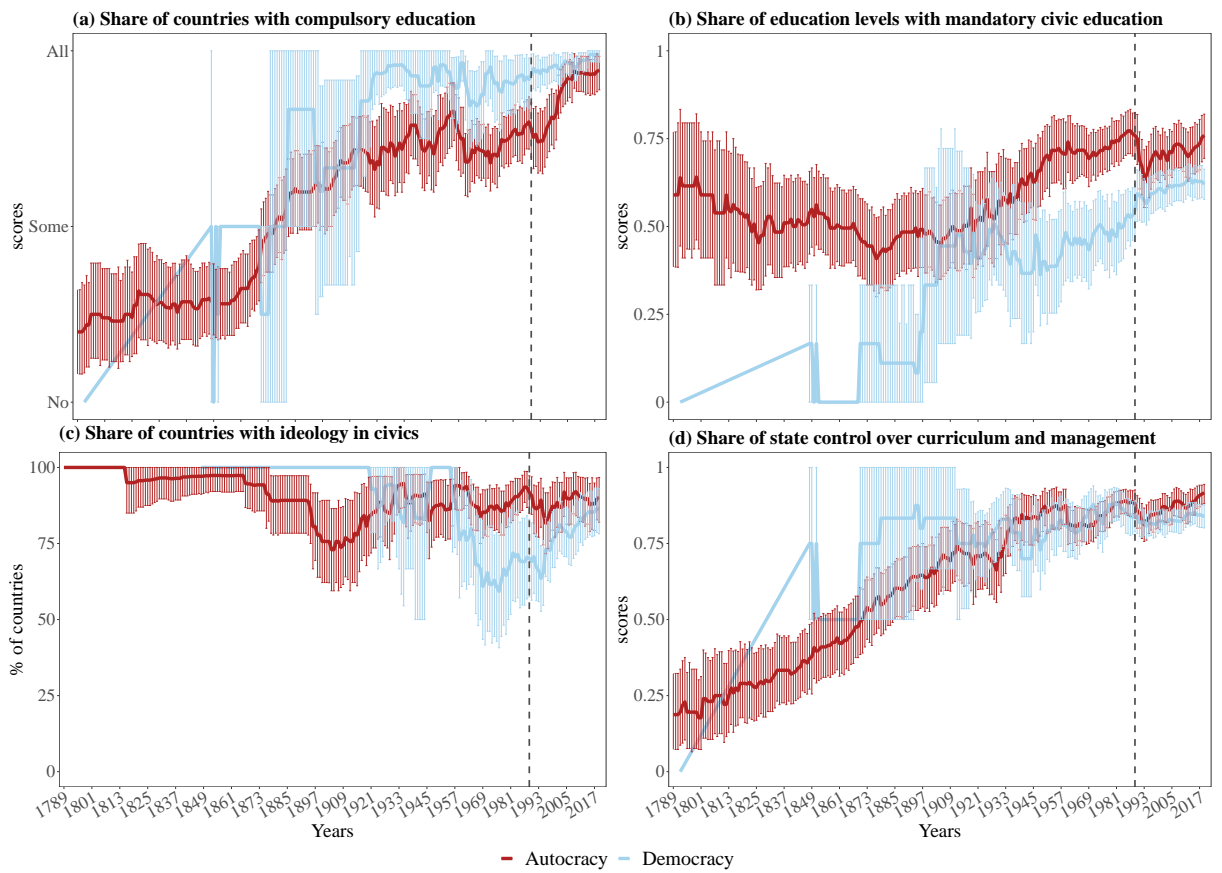


Figure SM.24. Patterns of education systems by regime type (using Lexical Index of Electoral Democracy to construct binary democracy measure)

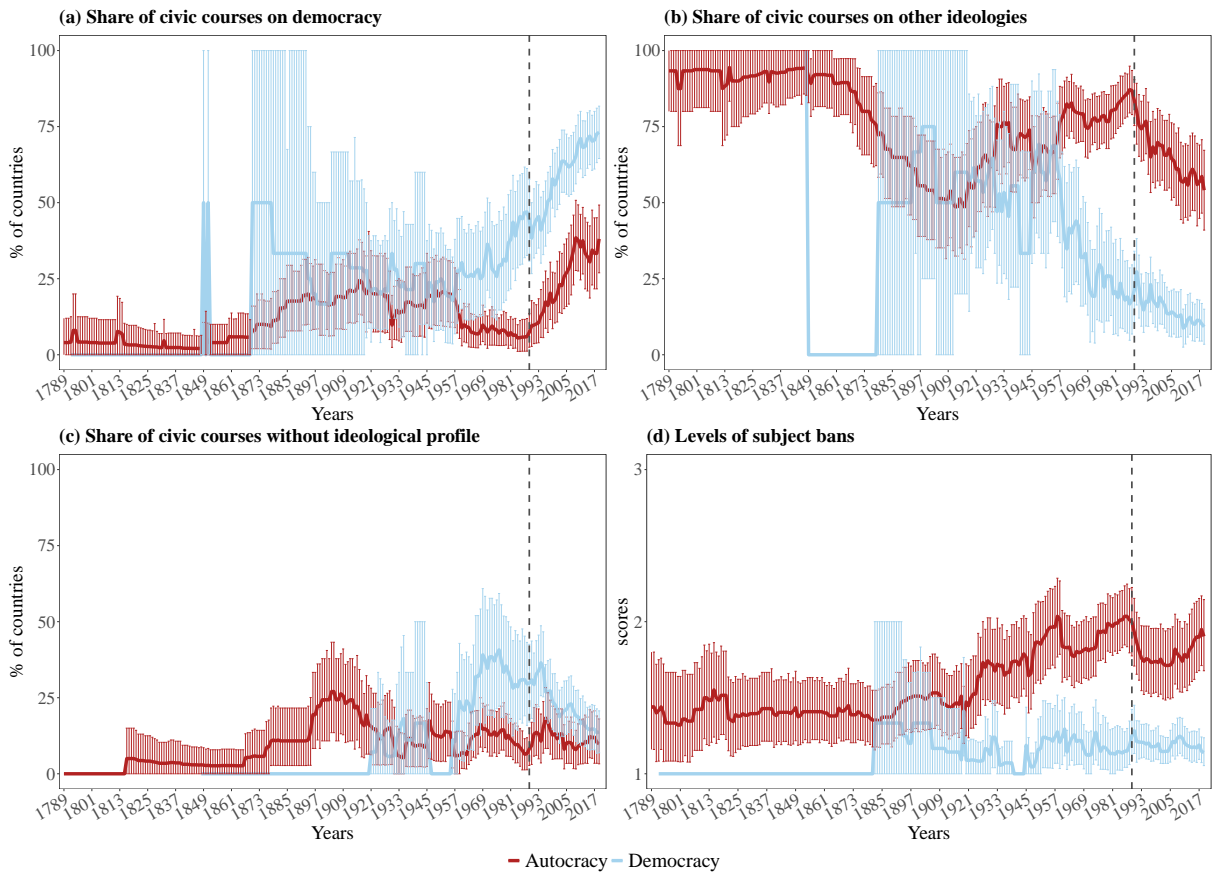


Figure SM.25. Patterns of education systems by regime type (using Lexical Index of Electoral Democracy to construct binary democracy measure).

Note: The score 1 on subject bans refers to the absence of explicit laws prohibiting topics or subjects in school. The score 2 denotes that one specific subject or topic is forbidden, while a 3-score denotes that several subjects or topics are forbidden.