H i C N Households in Conflict Network
The Institute of Development Studies - at the University of Sussex - Falmer - Brighton - BN1 9RE
www.hicn.org

Education and Conflict Evidence from a Policy Experiment in Indonesia

Dominic Rohner* and Alessandro Saia†

HiCN Working Paper 304

May 2019

Abstract: This paper studies the impact of school construction on the likelihood of conflict, drawing on a policy experiment in Indonesia, and collecting our own novel dataset on political violence for 289 districts in Indonesia over the period 1955-1994. We find that education has a strong, robust and quantitatively sizeable conflict-reducing impact. It is shown that the channels of transmission are both related to economic factors as well as to an increase in inter-religious trust and tolerance. Interestingly, while societal mechanisms are found to have an immediate impact, economic channels only gain importance after some years. We also show that school construction results in a shift away from violent means of expression (armed conflict) towards non-violent ones (peaceful protests).

Keywords: Education, Conflict, Civil War, Fighting, Schools, Returns to Education, Polarization, Protest

JEL classification: C23, D74, H52, I20, N45

Acknowledgements: We are grateful for helpful comments from Siwan Anderson, Chris Blattman, Michael Böhm, Patrick Francois, Luis Martinez, Munir Squires, Felipe Valencia Caicedo, and seminar audiences at Chinese University of Hong Kong, Hong Kong University, Simon Fraser University, Tilburg University, University of British Columbia, University of Chicago. Dominic Rohner and Alessandro Saia gratefully acknowledge financial support from the ERC Starting Grant POLICIES FOR PEACE-677595. We thank Esther Duflo for sharing her data with us.

First version: December 2018.

* Department of Economics, University of Lausanne and CEPR. E-mail: dominic.rohner@unil.ch

† Department of Economics, University of Lausanne. E-mail: alessandro.saia@unil.ch

"Education is, quite simply, peace-building by another name. It is the most effective form of defense spending there is." Kofi Annan, former Secretary-General of the United Nations

"No one is born hating another person because of the color of his skin, or his background, or his religion. People must learn to hate, and if they can learn to hate, they can be taught to love (...)." Nelson Mandela, Long Walk to Freedom

1 Introduction

Armed civil conflict is a major source of human suffering and obstacle to development, with political violence killing between 1945 and 1999 an estimated 16.2 million people in 127 civil wars (Fearon and Laitin (2003)), and an average civil war being estimated to lowering GDP by about 15 percent (Collier (2008)). While research on conflict has been thriving in recent years, in many articles the focus has rather been on the impact of exogenous economic, climatic or geological shocks than on public policies.[1] As discussed below, especially studies on the role of education are very scarce and are typically bound to cross-country correlations.

The very limited academic attention to the link between education and civil conflict contrasts sharply with the various anecdotal accounts, statements of peace negotiators and media reports praising education as a key long-run solution for curbing fighting.[2] This belief of journalists and diplomats expecting education to promote peace may also be shared by armed groups trying to perpetuate fighting and instability and which deliberately target schools in their attacks, such as Boko Haram (which loosely translates into "Western education is forbidden") in Nigeria or the Taliban groups in Pakistan and Afghanistan.[3]

The theoretical reasons for which one may expect education to be a rampart against civil conflict and political violence are manifold. As synthesized in De la Briere et al. (2017) (and shown formally in the underlying model of Rohner (2016)), turning physical wealth into human capital makes it harder to appropriate, and hence a less attractive "prize", and educated people have better job market prospects and thus higher opportunity costs of fighting instead of working. Moreover, schooling (if well designed) can transmit values of tolerance and open-mindedness, and education has been associated with more rational decision making (Kim et al. (2018)), raising potentially a person's awareness of the negative-sum nature of war. In contrast, education may also potentially foster social unrest and conflict, by increasing people's aspirations and means to engage in collective action against the regime in place. Further, when education is

[1] See e.g. the recent survey of Rohner (2017).

[2] See e.g. Inquirer, 3 February 2013, "Education is the lasting solution to Mindanao war"; Times of India, 22 March 2016, "When extremism stalks the students: Educational solutions to India's conflict zones".

[3] See e.g. BBC News, 24 November 2016, "Who are Nigeria's Boko Haram Islamist group?"; Reuters, 1st December 2017, "Nine killed as burqa-clad Taliban attack Pakistani college".

misused as mean of indoctrination it may also stir up inter-group tensions and boost nationalist sentiments. Which of these potential mechanisms has the strongest effect is to a large extent an empirical question that still awaits an answer. In the current paper we shall study several possible channels linking education to political stability.

In particular, we will in what follows carry out an empirical investigation of the impact of education on armed civil conflict intensity, exploiting a quasi-natural experiment in Indonesia. We study the impact of the INPRES Program which represents one of the largest and fastest school construction programs ever implemented. Between 1974 and 1978 over 61,000 new primary schools were built, amounting to more than doubling the stock of schools. The variation introduced by the INPRES program has first been exploited by Duflo (2001) to study labor market outcomes, and has since then been applied to other topics different from conflict, as discussed below.

While Indonesia is due to its size and social and economic heterogeneity an ideal country for studying the determinants of conflict, the striking scarcity of statistical studies on political violence in Indonesia may well be due to the lack of readily available conflict data ranging back far enough. In particular, existing measures of conflict only start after the end of the INPRES program implementation, ruling out any difference-in-difference analysis. To overcome this challenge, we have built our own novel and very extensive dataset of conflict events at the local district level (*Kabupate* in Indonesian), covering all of Indonesia over the period 1955-1994. Using techniques of web crawling and scraping and text recognition, we have drawn on information from over 820,000 newspaper pages to code variables of conflict events taking place in a given local district (*Kabupate*) and year. Our panel dataset contains 289 districts (*Kabupate*) over 40 years, resulting in 11,560 observations.

We carry out extensive sensitivity tests with respect to our novel conflict measure. Reassuringly, we find that for the years of overlap our coding of conflict and peace coincides in 86 percent of cases with the well-established conflict measure from GDELT (GDELT (2018)), which is higher than the level of correspondence between GDELT and other existing conflict data from ICEWS (ICEWS (2018)) or NVMS (NVMS (2019)) (for which there is no temporal overlap with our sample period).[4] Visual inspection (in the Appendix Figure 6) also highlights a parallel evolution of conflict events for our measure and the GDELT data. Concerns about reporting bias or other measurement errors are further attenuated by the fact that our point estimates are very stable when broadening the set of keywords. We also carry out an extensive Monte Carlo analysis of modifying the keywords used, and show that our results are robust for a variety of alternative algorithms, methods or news sources. Importantly, a part of our analysis does not require conflict data ranging back before the end of INPRES school construction, and for

[4]The National Violence Monitoring System (NVMS) data is used e.g. in the recent work of Bazzi and Gudgeon (2018) and Bazzi et al. (2019).

this analysis we are able to replicate our findings using alternative datasets, namely GDELT, ICEWS and NVMS. Reassuringly, our results go through when using these established datasets.

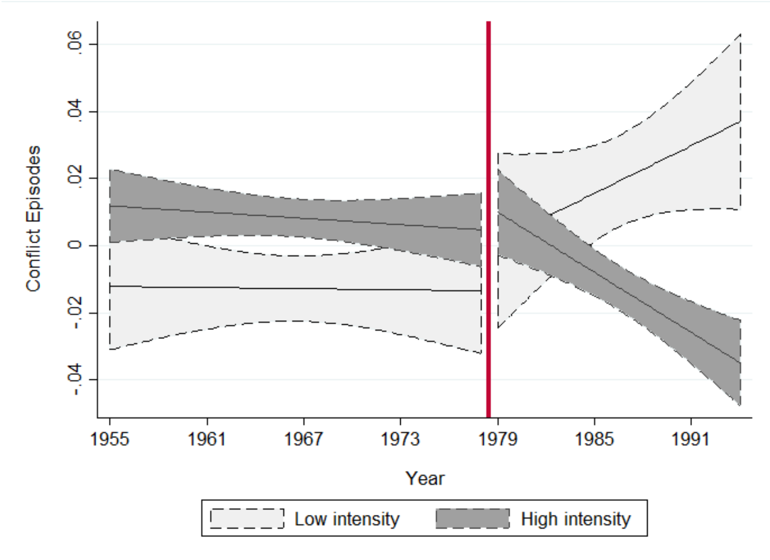
Our main identification strategy relies on a difference-in-difference approach, where we exploit the impact of *sharp* changes in education provision, resulting in *sharp* changes in conflict measures. As discussed in the previous literature (e.g. Duflo (2001)), and analysed more formally in the Appendix A.1, the goal of the INPRES program was to achieve a similar school density throughout the country, implying the construction of more schools in areas with initially fewer schools and a need to catch up. As shown below, the initial school density of a particular district is to a substantial part idiosyncratic.

As illustrated in Figure 1 below (and assessed in more detail in Appendix B.3), the pre-reform trend of conflict events is parallel in the areas with more versus less INPRES school construction. Following the introduction of the INPRES program there is a *sharp* opening of a gap between the conflict trend curves, with areas benefitting from heavy INPRES school construction being on a much more peaceful path. The gap between high- and low-INPRES construction areas then widens over time, exactly as one would expect due to the fact that (mechanically) the number of pupils benefitting from INPRES keeps rising until several years after the completion of INPRES school construction in 1979.[5] Note that while (as documented further below) the global fall of communism has contributed to a surge in political violence in Indonesia at the end of the 1980s and in the 1990s, the areas with fewer INPRES schools constructed were hit much more.

Despite these reassuring stylized facts on the common pre-trend, we put in place a variety of controls to filter out potential confounding factors. First of all, we include in all regressions district fixed effects (there are 289 districts) that control for time-invariant potential confounders such as e.g. remoteness or topography, and annual time effects (there are 40 years) accounting for global shocks (e.g. oil price shocks or elections). In addition, in our preferred specifications we replace the annual time dummies by province-year fixed effects, picking up any shocks taking place at the level of the 26 Indonesian provinces (think e.g. of the Indonesian government invading Eastern-Timor in 1975, or stepping-up repression in Aceh in 1990), and also include district-specific time trends, which filter out any long-run developments (e.g. less developed areas catching up to more advanced areas). In these most demanding specifications the only identifying variation comes from the *sharp* changes following the INPRES school construction start. Note that we also show robustness of results for controlling for major socio-economic measures interacted with post-INPRES dummies or time trends, for the other major government

[5]Children would typically attend newly constructed INPRES schools between the ages 7 and 12 (Duflo (2001)), which means that around 1984 the last "INPRES pupils" leave primary school and it may be more than 10 years later when parts of them finish university. Hence we expect increasing effects of INPRES on education (and hence conflict) until the end of our sample period (1994) with any potential ceiling being reached after 1994.

Figure 1: Intensity of INPRES school construction and conflict events



Source: Authors' computations from Duflo (2001) and own conflict data. Conflict data is obtained using the procedure described in Section 4.1. The figure shows the linear prediction plot with confidence intervals of normalized conflict events over pre-1978 and post-1978 periods (i.e., the year when INPRES program is completed). Normalized conflict events in a district-year are computed by removing the sample mean of conflict episodes observed in the whole sample in the corresponding year, as well as the district mean over time. Low intensity and high intensity indicate areas with more versus less INPRES school construction, respectively. Low (high) areas are defined as all districts where the number of schools is below the 25th percentile (above the 75th percentile).

construction program (water and sanitation), for migration, for weather or natural resource price shocks. We also find that our results are robust for alternative conflict data, for other econometric specifications, or when restricting the analysis to a series of subsamples. Finally, we further deepen the analysis, drawing on the synthetic control group approach, allowing to obtain a very close pre-treatment match for low- and high-INPRES districts. Our identifying assumption is that when filtering out potential confounders with the aforementioned battery of fixed effects and controls the intensity of INPRES exposure becomes a plausibly exogenous variable.

We find that indeed INPRES school construction has led to a statistically significant decrease in conflict, and that the magnitude of the effect increases over time (which was expected, given that mechanically the numbers of treated pupils at fighting age becomes larger over time). The effect is quantitatively sizable with a one standard deviation increase in schools built (1.25 more schools) resulting at the end of the sample period in a drop in three-quarters of the baseline conflict risk. Our results are robust to a vast array of robustness checks with respect to estimator, specification, measures, data construction and potential confounders. As far as heterogenous effects are concerned, schooling is found to matter both for areas with and without previous fighting and education reduces fighting across economic, ethno-religious and political types of conflict.

The analysis of channels of transmissions reveals that the conflict-reducing impact of schooling is both greater in districts with larger religious polarization and in those with higher economic returns to schooling. We also show that while societal facts immediately affect the impact of education and conflict, the economic channels of transmission only start to affect the conflict-reducing impact of education after some years. Drawing on individual data we further deepen the analysis of the role of religious tolerance as channel of transmission. We find that education boosts inter-religious trust and tolerance, as well as local community involvement. Interestingly, this effect is not driven by changing religious attainment, as we show that exposure to school construction does not affect religiosity. This implies that the increase in inter-religious tolerance cannot be mechanically attributed to lower religious observance. Interestingly, we also find that school construction only lowers violent means of resistance, but does not affect the propensity to engage in peaceful protests. Put differently, education makes people being not less but if anything more interested and willing to engage in local collective action, but pursuing their goals using a peaceful strategy of "voice" rather than "violence".

The remainder of the paper is organized as follows: Section 2 surveys the related literature and Section 3 provides an overview of the historical context. Section 4 introduces the data used and Section 5 is devoted to the presentation of the identification strategy and baseline results. Section 6 displays all robustness checks, Section 7 presents results on heterogeneous effects, and Section 8 investigates the underlying channels and mechanisms. Finally, Section 9 concludes. Additional results are relegated to the Appendix.

2 Literature Review

Closest and most relevant to the current paper is the empirical literature studying the effect of education on conflict. There exists cross-country evidence that education correlates negatively with conflict (Collier and Hoeffler (2004); Thyne (2006); Barakat and Urdal (2009); and Østby and Urdal (2011)). Yet higher education investments in a given country are not chosen at random and may correlate with a variety of confounding factors (such as e.g. other policies affecting conflict).[6] There exists to the best of our knowledge no paper yet that provides evidence from a (quasi-)natural experiment allowing for a causal identification of the impact of education on conflict. This is the gap in the literature that we seek to address in the current contribution.

Somewhat related is also the literature studying the impact of education on individual behavior. One of the punchlines is that schooling tends to reduce violent behavior. In particular, education lowers the individual crime propensity (e.g. Lochner and Moretti (2004)) and educational attainment correlates negatively with an individual's propensity to enlist in armed rebellion

[6] There is also evidence for the opposite direction of causality, with civil wars driving down human capital accumulation (Shemyakina (2011); Verwimp and Van Bavel (2013)), and international military rivalries fueling education investments (Aghion et al. (2019)).

(Humphreys and Weinstein (2008); Tezcür (2016)).[7] At the same time, it has been found that education boosts various forms of civic awakening and involvement, such as voter participation, acquiring of political knowledge, support for free speech and rejection of domestic violence and political authority (e.g. Dee (2004); Milligan et al. (2004); Glaeser et al. (2007); Wantchekon et al. (2014); Friedman et al. (2016)). The impact of education on political protests is ambiguous with Campante and Chor (2012) and Campante and Chor (2014) finding that education raises the willingness to participate to political protests – especially when more schooling is not matched by better employment opportunities – while Passarelli and Tabellini (2017) find that educated people less often participate to public demonstrations.

While it has also been found that education tends to reduce racism and increase inter-religious tolerance and the taste for cultural diversity (Hainmueller and Hiscox (2007); Roth and Sumarto (2015)), this channel is modulated by the educational content. Notably, the school curricula and teaching practices (such as copying from the board versus working on projects together) affect the level of government support and nationalism (Cantoni et al. (2017); Clots-Figueras and Masella (2013)), as well as student beliefs and their human and social capital (Algan et al. (2013); Cantoni and Yuchtman (2013)).

Last, but not least, our work is related to the literature using natural experiments, difference-in-difference estimations or randomized control trials (RCTs) to investigate the impact of education on topics other than conflict, such as health and fertility indicators (Osili and Long (2008); Somanathan (2008); Alsan and Cutler (2013); Breierova and Duflo (2004); Behrman (2015); Duflo et al. (2015)), labor market consequences (Duflo (2001); Duflo (2004); Akresh et al. (2018)), self-reported inter-group tolerance (Roth and Sumarto (2015)), bride price practices (Ashraf et al. (2019)), as well as local governance and public good provision (Martinez-Bravo (2017)). Out of the aforementioned papers, some draw on the same policy reform (INPRES) in Indonesia that we exploit in our current contribution (Duflo (2001); Duflo (2004); Breierova and Duflo (2004); Somanathan (2008); Roth and Sumarto (2015); Ashraf et al. (2019); Martinez-Bravo (2017); Akresh et al. (2018)). Yet all of these papers study phenomena that are very different from conflict.

In a nutshell, the current paper is the first study of the impact of education on conflict, drawing on exogenous variation in schooling. It creates a novel measure on conflict events in Indonesia and identifies channels of transmission through which education shapes the incentives for working versus fighting.

[7] There exists also evidence that the education levels of participants in Hezbollah, Hamas and Palestinian Islamic Jihad militant activities are –if anything– higher than those of the population average (see Krueger and Malečková (2003); Berrebi (2007)). Yet, given that the cells of these groups include typically a limited number of members, they may be able to pick the most skilled individuals out of a larger pool of applicants, which may be a driver of this correlation.

3 Historical Context

3.1 INPRES Program and Education in Indonesia

In 1973, the Indonesian government launched one of the most ambitious school construction programs ever enacted, both in terms of speed and scale. Between 1973-1974 and 1978-1979, more than 61,000 primary schools were built, more than doubling the number of schools in Indonesia. This led to an average of two schools constructed per 1,000 children aged 5 to 14 in 1971, with enrollment rates among children aged 7 to 12 increasing from 69 percent in 1973 to 83 percent by 1978 (Duflo (2001)). This program, labelled "Sekolah Dasar INPRES", was designed by the central government, mainly funded by oil revenues, and stipulated that the number of schooling places to be built had to be roughly proportional to school-aged children not enrolled before the program (Martinez-Bravo (2017)). The newly created schools were of similar size with each school being designed to host on average three teachers and 120 pupils of primary school age, which is normally between ages 7 and 12 in Indonesia. Importantly, efforts to train more teachers were stepped up in parallel, with the result that the share of teachers meeting the minimum qualification requirements did not significantly drop betweeen 1971 and 1978 (Duflo (2001)).

An important question is to what extent school construction in Indonesia actually boosted years of education achieved and employment opportunities. Interestingly, according to the 1971 Census (IPUMS (2018)) less than 37 percent of the population work in primary sector activities such as agriculture, fishing, forestry and mining, with the lion's share of employment being in the second and third sectors for which schooling typically matters substantially. Indeed, Duflo (2001) finds that each primary school built per 1000 children has resulted in an average increase of 0.12 to 0.19 years of education, and a wage increase of 1.5 to 2.7 percent. Accordingly, she has estimated economic returns to education to lie between 6.8 to 10.6 percent. These effects of school construction have been found to persist over time. According to Akresh et al. (2018), both men and women exposed to the program attain more education and reach higher living standards, while labor market effects are restricted to men. Akresh et al. (2018) also find that "these benefits are transmitted to the next generation. Children with fathers or mothers who were exposed to the school construction program obtain more education. Significant effects are observed at all levels of schooling beyond primary school, but the largest impacts are seen in tertiary education with effect sizes indicating a 20 to 25 percent increase in the likelihood of the second generation child completing university" (p. 43).

In terms of educational content of Indonesian primary schools, there is of course an obvious focus on basic literacy and mathematical skills. Still, as pointed out by Nishimura (1995) and Roth and Sumarto (2015), the Indonesian school curriculum has reserved some weekly hours on all education levels for the study of he principles of the state ideology *Pancasila*, which includes as main principles the belief in God allowing for freedom of religion / religious tolerance, humanitarianism, national unity, consultation as well as social justice. These principles have been kept vague and have been interpreted differently by different rulers, but what stands out in our context is the importance of advocating religious tolerance, which speaks to some of our findings, as discussed below.

Finally, it is important to note that the secular primary school sector –of which "Sekolah Dasar INPRES" is part– co-exists with the traditional islamic schools (the "madrasah", "langgar" and "pesantren"), with pupils being able to either fully opt for secular or islamic schooling or for attending both for some hours per day (see Postlethwaite and Thomas (2014)). An increased offer of nearby secular schools put in place by the INPRES program may lower the relative influence of islamic education, and this reduced relative importance of mono-religion schools may in turn increase the social interaction of pupils from different religions, resulting potentially in higher inter-religious tolerance.[8] Related to this, Bharati et al. (2017) show that the INPRES program has statistically significantly boosted public school attendance, while a negative, non-significant effect is found of INPRES school construction on private school attendance (which typically includes islamic education). Thus, while part of the effect of INPRES school construction may be due to more inter-religious interaction, the lion's share of the impact appears to be due to boosted educational attainment.

3.2 Conflict in Indonesia

A former Dutch colony, Indonesia has won independence in 1949.[9] Its form has kept evolving, with the Western segment of New Guinea being officially recognized as part of Indonesia in 1969 by the United Nations, and the former Portuguese territory of East Timor (Timor-Leste) belonging to Indonesia from 1976 to 2002. After a period of unruly parliamentary democracy, President Sukarno declared in 1957 martial law and introduced "Guided Democracy". After a failed coup, his power faded, and from 1967 until 1998, the country was ruled by the authoritarian regime of Suharto, and is again a democracy since his demise.

Indonesia is located in an archipelago containing 13,466 islands (of which 922 are inhabited), counts today roughly 260 million people and is characterised by very rich ethnic diversity (with the largest groups being the Javanese 40.1%, Sundanese 15.5%, Malay 3.7%, Batak 3.6%, and Madurese 3% (CIA (2018)). There is also a fair amount of linguistic and religious heterogeneity

[8] For example, Merlino et al. (2019) find for the US that more interracial contact during childhood boosts inter-racial relationships and tolerance later in life.

[9] This subsection builds on the accounts in Brown (2003); CIA (2018); BBC (2018); and Encycl. Britannica (2018).

with over 700 languages used and most major religions being present (Muslim 87.2%, Protestant 7%, Roman Catholic 2.9%, Hindu 1.7%, according to CIA (2018)).

Indonesia has during our sample period 1955-1994 suffered from a substantial amount of conflict, with its sources and reasons being as heterogenous as the country itself, and where ethnic and religious cleavages and the scattered archipelago geography may well have played important roles. Part of conflict events were driven by secessionism, such as separatist rebellion in Aceh being present since 1953, the rebellion in Western Sumatra and North Sulawesi (Sulawesi Utara; North Celebes) in the second half of the 1950s, the Darul Islam movement in West Java in the 1950s and early 1960s, the separatist Free Papua Movement since the early 1960s, armed Maluku secessionism in the 1950s and 1960s, or the armed resistance of East Timorese during the period of incorporation into Indonesia following the invasion in 1975. Other political violence was at least partly motivated by ideology, such as the anti-communist purges following the failed 1965 coup. Moreover, some fighting has been linked to religion, with the (rather secular) Indonesian governments at various moments clashing with Muslim parties and movements, leading e.g. to the ban of the Masyumi party in 1960 or the riots in Tanjung Priok in 1984. Finally, many instances of localized communal or ethnic rioting and fighting have taken place throughout our sample period.

As displayed graphically in the Appendix Figure 6, there has been a surge in conflict events in the early 1990s. After the Fall of Berlin Wall in 1989 and the collapse of communism, the domestic and international public opinion started paying increasing attention to President Suharto's dismal human rights record and this increasing pressure triggered rising repression and purges. Simultaneously, in line with the worldwide unfolding of secessions and state break-ups during this period (e.g. in Eastern Europe and the Balkans), the rising tensions in Indonesia of the early 1990s were particularly striking in provinces with secessionist movements such as e.g. Aceh (see the Online Appendix Figure B3). Note that a crucial robustness analysis carried out in Online Appendix B.12 shows that our results do not hinge on the inclusion of the troubled 1990s or of any particular Indonesian province.

4 Data

4.1 Construction of Conflict Variables from Newspaper Data

As described in more detail in Online Appendix B.1, the lack of suitable existing conflict measures spanning over the time period of our sample has led us to engage in data collection and in the construction of a novel conflict measure. Our approach to construct a novel geo-referenced dataset of conflict-related events in Indonesia consists of five steps.

The *first step* was to find a high-quality newspaper that is digitally available and covers a long enough time period. As discussed in more detail in Online Appendix B.1, the Sydney Morning Herald (thereafter, SMH) fulfills all our requirements, and has the advantage of being Australian – a country with traditionally quite detailed news coverage of its neighbor country Indonesia. Founded in 1831, the SMH is the oldest continuously published newspaper in Australia and currently has a readership of roughly half a million people (Morgan (2018)). As discussed in Appendix B.1, it is unlikely to suffer from any particular bias, and its digital archive allows us to construct a database of violent events in Indonesia between 1955 and 1994. As shown below, our results are very similar if we use an alternative media source, i.e. the Canberra Times – an other serious, yet smaller news outlet.

After having identified the newspaper, the *second step* was to perform a first selection of the articles related to Indonesia. In particular, we searched over 820,000 articles available in SMH archive and downloaded all those containing at least once the word "Indonesia" (the resulting set of articles was of around 34,000).

In a *third step*, we used natural language processing algorithms to analyse the content of all articles, storing all sentences where at least one conflict related term was present.[10] Finally, in the (*fourth and fifth steps*), we started out using a Named Entity Recognition algorithm to identify all locations referred to, and then matched locations to geo-coordinates.[11].

When confronting our conflict measure drawn from SMH to the existing conflict variable of GDELT (2018) we find, as discussed in Appendix A.9, that in 86 percent of cases our conflict variable takes the same values as GDELT (with which there is a temporal overlap for 1979-1994). Importantly, also visual inspection of Appendix Figure 6 confirms the parallel evolution of our measure and GDELT for the years of overlap.[12]

As discussed below and explained in much detail in the Online Appendix B.1, we have performed a wide set of robustness exercises to assess the validity of our conflict measure, i) focusing on an alternative newspaper (the Canberra Times), ii) using an alternative python algorithm to identify locations, iii) applying alternative matching scores for geo-identification, and finally, iv) exploiting three alternative conflict databases (covering a shorter time period) to replicate our analysis (GDELT (2018), ICEWS (2018), and NVMS (2019)). Our results are found to be

[10]In the main analysis we focused on: "conflict" "battle" "assault" "kill" "riot" "attack" "turmoil" "unrest" "warfare" "solider" "army" "insurgent" "terrorist" "disorder" "revolt" "massacre" "strike" plus all their variations. We also performed extensive robustness checks narrowing or widening the set of terms.

[11]For illustration, fighting events captured include e.g. the ones mentioned in the following newspaper sentences: "Fierce fighting was reported to have broken out last Wednesday in Macassar on the island of Celebes after rebels attacked an army patrol" (30/12/1957); "Indonesia has admitted that unrest occurred near Manokwari recently, and that troops were used to quell the trouble" (28/08/1965); or "The riots were confined to one area of Tanjong Priok, a densely populated and predominantly poor suburb of northern Jakarta, close to the port" (15/09/1984)

[12]Note that we cannot directly confront our data to ICEWS (2018) and NVMS (2019), as the SMH is only available until 1994, while ICEWS starts in 1995, and NVMS (partial) coverage begins in 1998.

robust to these alternative ways of constructing the conflict measure.

4.2 Other Data Used

Our education measures are based on the number of schools constructed between 1973-1974 and 1978-1979 per district by the INPRES program – as described above in section 3. In particular, our main education variable is given by the number of schools constructed under the auspices of the INPRES program per 1000 children of primary school age in a given district in 1971. The raw data on this is taken from Duflo (2001).

Further, the variables of the index of religious polarization, returns to schooling and other district-level variables have been constructed using the 1971 Population Census conducted by the Central Bureau of Statistics of the Republic of Indonesia (IPUMS (2018)). We focus on 1971, as this is prior to the INPRES intervention. For constructing the polarization measure at the district level we apply the polarization formula described in Montalvo and Reynal-Querol (2005) to the religion share data of the 1971 census.

Individual-level variables on religious tolerance and community participation were retrieved from the 5th wave of the Indonesian Family Life Survey (IFLS) (Strauss et al. (2016)). Finally, district borders for Indonesia at the time of the INPRES program were obtained from the Digital Atlas of Indonesian History (Cribb (2010)).

4.3 Descriptive summary statistics

Table 1 displays district-level descriptive statistics for the explanatory variable and the dependent variables (further descriptive statistics for the other variables are presented in Online Appendix B.2). Our sample consists in slightly less than 300 districts over 40 years. On average, 2.35 schools per 1,000 children aged 5 to 14 were constructed under the INPRES program in less than six years. While all districts have been exposed to the program, the intensity of new school constructions varies widely across regions. Further, using data from our preferred source (i.e. *Sydney Morning Herald*), the probability of observing at least one conflict in a given district and a given year is around 8%. This number increases to 13% when we use a wider set of conflict-related terms (see Appendix A.8). The conflict probability is lower when we consider another, smaller newspaper (i.e. *Canberra Times*). The lower conflict probability obtained with this second data-source is consistent with its incomplete coverage of our time period of interest. Finally, the last line reports event probabilities obtained when we combine the two sources.

SOURCE: School construction data from Duflo (2001). Conflict data is obtained using the procedure described in Section 4.1.

Table 1: Descriptive statistics

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
(# Schools / # Children)$_i$	2.35	1.257	0.591	8.598	289
Conflict Episode$_{it}$:					
Sydney Morning Herald [SMH]	0.077	0.267	0	1	11,560
SMH - Broader Definition	0.13	0.336	0	1	11,560
Canberra Times [CB]	0.06	0.238	0	1	11,560
SMH and CB Combined	0.105	0.307	0	1	11,560

5 Empirical Strategy and Main Results

5.1 Identification strategy

INPRES has been widely acknowledged as one of the fastest and largest-scale school construction program worldwide to date; and when it started in 1973 it boosted education much more in some districts than in others. In our analysis below we shall exploit the differential increase in education across districts and study whether more intense school construction under INPRES has led to a relative decline of conflict with respect to other districts less affected by INPRES.

This framework implies a series of econometric challenges:

Different underlying conflict risk at district level – Districts with initially fewer schools may have a propensity for conflict that is generally higher or lower. Such level effects are picked up by the fact that we include district (*kabupate*) fixed effects.

National or regional policies – A potential confounder could be other policies or programs taking place at the same time as INPRES. To address such concerns, any general nationwide policies are filtered out by annual time effects. Further, we replace in most specifications the general annual time effects with province-specific annual time dummies (i.e. we include Province * Year fixed effects). In a robustness check we also control for the intensity of exposure to the water and sanitation program which took place simultaneously.

Different time trends for districts – Another concern could be that places with initially fewer schools may have another pre-trend or may catch up with more urban neighborhoods independently of the INPRES program. First of all, as far as a common pre-trend is concerned, as illustrated in the introduction in Figure 1, and shown in more detail in the Online Appendix B.3, before the start of the INPRES program in 1973 there was a common trend of conflict events in districts with a high initial school density (where few additional schools were needed under INPRES) compared to districts with initially only few schools (and where INPRES heavily engaged in school construction to meet the target of homogenous school density nationwide). Most importantly, we further include in many specifications linear district time trends, which would pick up any mechanical convergence or divergence effects of different types of districts. Note also that we have an extensive robustness section, where the synthetic control group approach is exploited for constructing identical pre-trends.

Different shocks in districts – Similarly, one could worry about economic or non-economic shocks hitting different areas differently in the same years of INPRES implementation. Imagine for example that by accident places with more INPRES school construction are affected differentially by some economic shock hitting Indonesia during the same period. First of all, in a robustness table we interact the post-reform years with the initial pre-reform school enrollment rates in 1971. This is an important robustness check, as the planing of the number of schools to build was based on these 1971 enrollment rates, and when conditioning on these, the intensity of INPRES school construction becomes a random variable (see Appendices A.1 and A.2). Hence, differential shocks in places with more INPRES school construction would have to be due to pure coincidence. We also interact in robustness checks the post-INPRES period with a battery of other socio-economic covariates from the 1971 census. Further, we take into account the increasing effect of school construction over time (given that the total number of additional schooling years enabled by INPRES increases over time, as discussed above). Hence, in order to confound the impact of INPRES, such other potential shocks would also have to show the same inter-temporal pattern of intensification.

5.2 Correlates of school construction and pre-trend

The goal of the INPRES school construction program has been to achieve a homogenous level of school enrollment across Indonesia, and indeed the intensity of school construction under INPRES in a given district has been claimed to be essentially driven by the pre-INPRES school enrollment rate of school-aged children (in places with initially too few schools, more schools were built; see Duflo (2001)). In order to check this formally for our data, we carry out in Appendix A.1 an analysis of the determinants of school construction. In particular, we regress the number of schools built under INPRES on pre-INPRES enrollment rates of school-aged children and a variety of pre-INPRES socio-economic covariates. We find in Table A1, as expected, that INPRES school construction is only determined by pre-INPRES enrollment rates of school-aged children and unrelated to a wide array of socio-economic variables. Note that we also show below that all our results go through when controlling for pre-INPRES enrollment, as well as a variety of socio-economic factors (see Appendix A.2).

While the fact that INPRES school construction does not correlate with a series of potential confounders is reassuring, we can also directly check the common pre-trend assumption. This is what we have illustrated in Figure in Section 1, and study in more depth in Online Appendix B.3. Slicing the sample in a variety of ways between districts with more versus less intensive school construction, we conclude that the common pre-trend assumption appears reasonable. Notice that we also carry out below in section 6.1 an extensive synthetic control group analysis where by construction the synthetic control group has a parallel pre-trend. Our results are robust to the synthetic control group analysis.

5.3 Econometric specification

In terms of the variable construction, our unit of observation is the district-year and for our empirical analysis we combine district-year-level violence data with district-level data on the number of new schools built between 1973-1974 and 1978-1979 (district-level data from Duflo (2001)). The dataset covers 289 districts (*Kabupate* in Indonesian) across 26 provinces over the period 1955-1994.

Applying the logic of difference-in-difference settings, we exploit both the variation over-time (i.e. difference pre/post) and over-space (i.e. difference in the intensity of the programme across regions). We start with the standard specification

$$Conflict_{it} = \alpha + \beta_1 \frac{\#Schools\ Built}{\#Children}_i + \beta_2 Post - 1978_t + \beta_3 \frac{\#Schools\ Built}{\#Children}_i * Post\ 1978_t + \epsilon_{it},$$

where the variable $Conflict_{it}$ is a dummy that takes a value of 1 if a violent event was observed in district i in year t. The variable $(\#Schools\ Built/\#Children)_i$ represents the number of primary schools constructed under the INPRES program.

The dummy $Post - 1978_t$ takes a value of 1 for the first year when we expect the program to deploy major effects, as 1978 is the year when school construction is complete and it is also roughly the first year when the pupils first enrolling in the program 5 years earlier would be old enough to engage in violent activities (e.g. the report of Refworld (2001) states that "Indonesia's troubled provinces are said to use child soldiers as young as 12").

Further, the specification includes the interaction of $(\#Schools\ Built/\#Children)_i$ and *Post-1978*$_t$. This interaction term is our variable of interest, as we expect school construction under INPRES to deplete effects after 1978, and the more so the more schools were built under the program.

We include in all specifications district fixed effects (FE_i) and year fixed effects (FE_t). This means that both β_1 and β_2 will be absorbed by our fixed effects. Thus, the first baseline specification of column 1 of Table 2 that we estimate becomes

$$Conflict_{it} = \alpha + \beta \frac{\#Schools\ Built}{\#Children}_i * Post - 1978_t + FE_i + FE_t + \epsilon_{it}.$$

Standard errors are clustered at the level of the 289 districts in all regressions (unless indicated otherwise). Note that in addition in column 2 of Table 2 we include a vector of district-specific linear time trends, while in column 3 of Table 2 we include both a vector of district-specific time trends as well as Province times year fixed effects (FE_{pt}).

The aforementioned specifications correspond to the simplest difference-in-difference design that only distinguishes between years before the treatment deploys effects (i.e. pre-1978) versus the years where the treatment is active (i.e. post-1978). This specification has the virtue of simplicity, but it does not take into account that the number of children treated by INPRES and reaching an age where they could get possibly enrolled in violent activities is increasing mechanically every year. As discussed above, typically the age where involvement in violent activities becomes conceivable in Indonesia is about 12 years (see Refworld (2001)).

Put differently, while INPRES starts to deploy effects from 1978 onwards (which is when the first INPRES intake reaches "fighting age"), we expect its impact to become larger every year, both in terms of the extensive margin (i.e, the number of INPRES pupils reaching potential "fighting age" increases), as well as in terms of the intensive margin (i.e. while enrollment in violent activities is conceivable at the age of 12, it becomes more likely in later teenage years).[13] For this reason we focus in columns 4-6 on a specification allowing for an increase in the treatment effect over time, by interacting the variable $(\#Schools\ Built/\#Children)_i$ with a variable defined as *Numbers of years since 1978* (i.e. a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on). In particular, in column 4 of Table 2 we estimate the equation

$$Conflict_{it} = \alpha + \beta \frac{\#Schools\ Built}{\#Children}_i * Years\ since\ 1978_t + FE_i + FE_t + \epsilon_{it},$$

and in columns 5 and 6 of Table 2 we again add a vector of district-specific time trends as well as Province times year fixed effects (FE_{pt}). Note that the variable of *Years since 1978*$_t$ will be

[13] Actually, one may expect the effect to stop increasing further after some period of time, which is indeed what we see in the early 2000s when using GDELT data (results available upon request). Given that the sample of our baseline analysis stops in 1994, unsurprisingly for our sample period the effect ever increases.

absorbed by the vector of time fixed effects.

While this second specification of columns 4-6 accounts for the increasing treatment intensity over time, it imposes a linear increase in treatment. Hence, in order to allow for a non-linear change in treatment effects over time we shall in columns 7-9 of Table 2 perform the initial difference-in-difference specifications of columns 1-3, but distinguishing three time windows for the period after 1978: i) [1979-1984], ii) [1985-1989] and iii) [1990-1994].

5.4 Baseline Results

Table 2 reports the baseline results of the specifications mentioned above. In the simplest difference-in-difference specification of columns 1-3 in all columns the coefficient of interest has the expected, negative sign and is statistically significant.

Focusing on our preferred, most demanding specification of column 3, we can see that the conflict-reducing impact of education is quantitatively substantial by any standards. As depicted in Table 2, the mean conflict likelihood in a given district year is 0.08, and the mean numbers of INPRES schools built per 1000 school-aged children is 2.35. As shown in column 3, building one more school per 1000 school-aged children reduces the conflict likelihood by almost -0.02, which is a quarter of the baseline conflict risk. Expressed in standard deviations, one standard deviation change in school construction (around 1.25 schools per 1,000 children) leads to a 8% standard deviations lower conflict risk every year after 1978.

As shown in columns 4-6, there is indeed evidence for the expected increasing effect of INPRES school construction over time. The coefficient of interest is of expected sign and statistically significant in all columns, and is quantitatively very sizable. The quantitative impact becomes even more impressive when taking into account the increasing effect over time. For example, consider for our preferred specification of column 6 the impact of one standard deviation greater INPRES school construction (i.e. around 1.25 schools more per 1,000 children) on the conflict likelihood in the end-of-sample year 1994, which is 16 years after its completion. This effect of INPRES amounts to roughly -0.06 [=-0.003*16*1.25], which corresponds to a decline in three quarters of the baseline risk of conflict (0.08), or, put differently, to a decline of 0.23 standard deviations of conflict risk.

Moving to columns 7-9, we can see that indeed the impact of INPRES school construction gets larger over time, with again the impact of school construction in the third period [1990-1994] amounting in our preferred column 9 to more than half of the baseline risk of conflict (0.08).

Table 2: Baseline results of the impact of INPRES school construction on conflict

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(# Schools / # Children)$_i$ * *Post-1978*$_t$	-0.0127***	-0.0101*	-0.0173***						
	(0.00448)	(0.00573)	(0.00610)						
(# Schools / # Children)$_i$ * *Years since 1978*$_t$				-0.00146***	-0.00175***	-0.00305***			
				(0.000421)	(0.000654)	(0.000705)			
(# Schools / # Children)$_i$ * *Years 1979-1984*$_t$							-0.00658	-0.00952*	-0.0168***
							(0.00427)	(0.00536)	(0.00591)
(# Schools / # Children)$_i$ * *Years 1985-1989*$_t$							-0.0111**	-0.0151*	-0.0253***
							(0.00540)	(0.00813)	(0.00837)
(# Schools / # Children)$_i$ * *Years 1990-1994*$_t$							-0.0218***	-0.0268***	-0.0489***
							(0.00594)	(0.0101)	(0.0110)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.450	0.506	0.346	0.450	0.506	0.346	0.450	0.507
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Province x Year FEs	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.08	.08	.08	.08	.08	.08	.08	.08	.08

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978*$_t$ takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978*$_t$ is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Years 1979-1984*$_t$ is a dummy taking a value of 1 for the years 1979-1984 (it is analogous for the two variables referring to the period 1985-1989 and 1990-1994, respectively). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Robustness Analysis

Below are described the various robustness checks carried out to assess the sensitivity of the main results displayed above in Table 2. In the current section we shall limit ourselves to a short account of the robustness analysis, with most of the robustness tables and further details being relegated to the Appendix.

6.1 Synthetic Control Method

Above we have shown in Figure 1 (and in more detail in Appendix B.3) that the common pre-trend assumption is supported by the data. Still, to go one step further, and make sure that the pre-reform trend of conflict events is indeed always parallel in the areas with more versus less INPRES school construction, we apply a transparent method of choosing counterfactual units: the synthetic control method.

In recent applications, the synthetic control method has proven to be a valid tool to assess the impact of policy-related events (see e.g. Abadie and Gardeazabal (2003), Abadie et al. (2010), Billmeier and Nannicini (2013), Saia (2017)) where i) it is possible to distinguish treated from untreated units and ii) the outcome of interest was a continuous variable (e.g. GDP, trade flows, ect). Unlike in most of the previous works, in our setting the outcome of interest is a dichotomous variable and, as explained in the previous section, our treatment of interest is the intensity of the school construction program (since the program was implement across the entire country). In order to employ the synthetic matching in our setting, we need to depart from previous works along two dimensions. Firstly, we define treated and control units based on the intensity of the treatment. That is, the potential counterfactual units for a given district are all districts where the intensity of the INPRES program was lower than for the unit of interest. Secondly, due to computational limitations (mainly related to converging problems of the Synthetic Control Method algorithm with dichotomous variables) we use as outcome of interest the number of conflict events observed in a 5-year time-window.[14]

For each district, we apply the synthetic algorithm to construct a counterfactual unit as a weighted combination of a group of potential counterfactual units.[15] Weights are selected in order to approximate the incidence of conflict events of the unit in question prior to the

[14] In other words, we divide our panel into 8 sub-periods and we collapse all units along this dimension. The 8 sub-periods are: [1955-1959], [1960-1964], [1965-1969], [1970-1974], [1975-1978], [1979-1984], [1985-1989] and [1990-1994]. The first 5 time-windows correspond to the pre-INPRES period.

[15] We implement the synthetic control method for all districts where the number of schools is above the $25th$ percentile. This is due to the fact that the synthetic algorithm requires a certain number of potential counterfactual units. Including all low-INPRES intensity units would create computational problems related to the very limited number of potential counterfactual units available (e.g. the district with the lowest number of INPRES schools would not have any potential counterfactual unit to be compared with).

implementation of the INPRES program, using a transparent data-driven procedure. The idea behind this method is that if the matching window is large enough, the weighted combination is able to replicate the structural parameters of our district of interest and successfully reproduce all the observed and unobserved determinants of conflict for the district in question. To ensure that the results are not driven by the inclusion of any particular district, we replicate this procedure using 500 different groups of potential counterfactuals, where each counterfactual group is computed randomly by drawing on two-thirds of all control districts.

In order to assess the total effect of the INPRES program at the national level we aggregate all *treated* districts and the corresponding synthetic counterfactual observations. In doing so, we are able to compare the actual incidence of violence observed in Indonesia with the distribution of violence observed in the 500 aggregate synthetic counterfactual units. We remove all districts for which the synthetic algorithm fails to provide a good match during the matching window.[16]

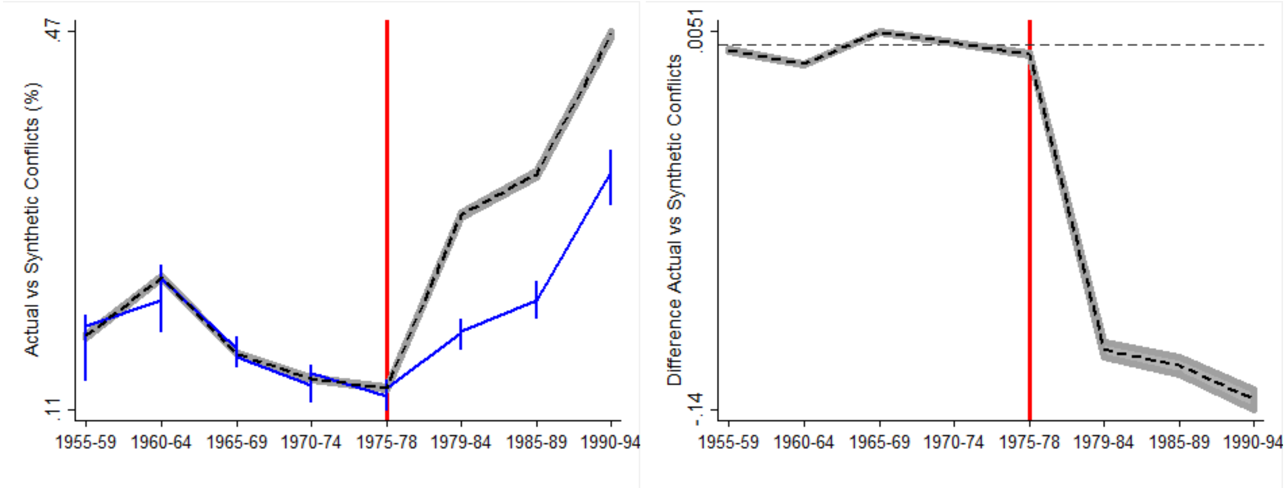
The left panel in Figure 2 plots the actual evolution of conflict events observed in our sample (solid line) and the one provided by aggregating synthetic units over the period of interest (dashed line). In the case of the pre-INPRES period, the synthetic counterfactuals provide a good approximation of the aggregate level of conflict events observed in Indonesia, and the synthetic (dashed line) and actual violence (solid line) behave very similarly. After 1978, the dashed line shows how violence would have developed if fewer schools had been constructed in each district. The two lines start to diverge substantially right after the end of the INPRES program and we can see that there are fewer conflict events in the districts with more INPRES schools compared to the synthetic counterfactual. The right panel in Figure 2 displays the evolution of the difference between actual and synthetic units.

6.2 Robustness to controlling for the impact over time of socio-economic district characteristics

If the last subsection was targeted at addressing concerns about the common pre-trend assumption, there may remain worries about particular shocks hitting after 1978 (by coincidence) the kinds of districts mostly affected by INPRES school construction. Given that we also document the dynamic increase of the impact of school construction over time, potential confounders would need to be characterized by the same dynamic profile of increasing effects, which further restricts the kinds of confounders that could affect our results. Still, below we shall focus on an array of socio-economic variables and interact them with the "post-1978", resp. "years since 1978" variables, which would pick up shocks related to particular socio-economic characteristics.

[16]In particular, we remove all units for which the difference between actual and synthetic aggregate observations during the pre-INPRES period is greater than $2 \cdot \sigma$ (conflicts events$_{pre-INPRES}$), where σ is the standard deviation.

Figure 2: Results with Synthetic Control Method



A) Levels B) Difference

Note - Left Panel: The solid line corresponds to the actual average incidence of violence observed in all districts, while the dashed line captures the average incidence of violence obtained from synthetic counterfactuals. The dark grey area around the dashed line indicates the 99% confidence interval. Each synthetic unit was computed as a weighted average of randomly drawn group districts where the intensity of the INPRES program was lower than in the district of interest. Weights are selected according to the incidence of conflict events of the unit in question prior to the implementation of the INPRES program. We remove all districts for which the synthetic algorithm fails to provide a good match during the matching window (see additional details in the text). Right Panel: The dashed line represent the average difference between actual incidence of violence observed in a district and the incidence of violence obtained from the synthetic counterfactuals of the Left Panel.

All results are displayed in Appendix A.2. Most importantly, we start in Table A2 with the enrollment rate of school-aged children in 1971, which in the analysis of Appendix A.1 was the most powerful predictor of school construction under INPRES (i.e. more schools were built in places with initially fewer schools). After that we investigate the impact of a wide set of other socio-economic indicators. We find that our results are robust to the inclusion of such additional control variables.

6.3 Robustness to controlling for the water and sanitation investment program

Finally, in Appendix A.3 we also control for the incidence of water and sanitation program being implemented at a similar point of time as the INPRES school construction program. This is an important robustness check, as first of all this program could represent a confounding factor and, secondly, the water and sanitation program was the second biggest large-scale investment program of the central government in this period (Akresh et al. (2018)), and hence any potential (mechanical) bias –due e.g. to reporting bias– should typically also be present for this second major investment program. Reassuringly we find that the estimated impact of school construction is unaltered when adding this further control, and we don't detect any impact on conflict of the water and sanitation program.

6.4 Robustness to controlling for migration

Another potential worry could be biases arising from migration. If migration is uniform and similar across districts and over time, it merely results in attenuation bias, making us underestimate the true effect of school construction. More worrying would be a situation where migration levels are large and potentially correlated with district characteristics. A priori, we expect this not to be a major issue. Duflo (2004) concludes that migration levels are not very large and does not detect any biases linked to selective migration in the context of her study. Akresh et al. (2018) find that the INPRES school construction program did affect migration flows, but only to a quite small extent. In particular, they find that "the school construction program increases migration rates by 0.7 and 0.8 percentage points respectively (...), [that] the increase in migration is concentrated in shorter distance moves within—rather than between—provinces" (p. 16), and that "the school construction program does not increase the share of people living in urban areas. They do appear to move to more valuable and larger housing" (p.18).

Nevertheless, we control for migration and urbanization patterns. In particular, we use Indonesian census data to construct the share of population in a given district having immigrated from another province, as well as the share of the population living in rural areas. The data construction is discussed in detail in Appendix A.4. As shown in Table A4 in Appendix A.4, the inclusion of these controls does not affect our results.

6.5 Robustness to climate and oil rents shocks

An important other source of potential confounders to consider are climate and natural resource shocks. Hence, in Appendix A.5 we investigate whether climate shocks such as temperature or rainfall variation could have been confounding factors for explaining the observed levels of conflict. We similarly study the role of oil world price shocks in extraction areas (following an identification strategy akin to Berman et al. (2017)). For all the variants of such shocks studied, we find the impact of school construction to remain very stable and statistically significant.

6.6 Robustness to alternative econometric choices and specifications

The fact that we cluster the standard errors at the level of treatment implies a conservative assessment of statistical significance. To assess the sensitivity of our statistical inference, in Online Appendix Section B.4, we display the significance levels for alternative levels of clustering. In particular, Table B3 allows for standard errors to be clustered at the level of the 26 Indonesian provinces (although this number of clusters is arguably below the conventional minimum levels), while Table B4 allows for standard errors to be two-way clustered at the district and year levels. In both cases the coefficients of interest remain statistically significant.

To further address remain concerns about inflated statistical inference we alter below the level of aggregation. In particular, as advocated by Bertrand et al. (2004), below in Table 3 as alternative specification we collapse the time dimension into "pre-" versus "post-" treatment. This specification does not allow to study the increase of treatment effects over time, which is the main reason why it is only used as robustness check and not as main specification. Reassuringly, this very different econometric specification yields very similar results as our main specification.

Table 3: Robustness: Alternative econometric specification

	Dummy Conflict$_{iT}$		(log) Years with Conflict$_{iT}$	
Dep. Variable:	(1)	(2)	(3)	(4)
(# Schools / # Children)$_i$	-0.0866***	-0.0942***	-0.0938***	-0.0904***
	(0.0151)	(0.0215)	(0.0214)	(0.0291)
Conflict Prior to INPRES Program:				
Dummy Conflict$_{iT-1}$	0.458***	0.395***		
	(0.0512)	(0.0596)		
(log) Years with Conflict$_{iT-1}$			0.620***	0.581***
			(0.0532)	(0.0610)
Observations	289	289	289	289
R-squared	0.297	0.382	0.487	0.540
Province FEs	No	Yes	No	Yes
Sample Mean	.37	.37	.49	.49

NOTE: The unit of observation is a district i in period T, where T represents the period [1979-1994], and $T-1$ corresponds to the period [1955-1978]. LPM estimates are reported in the first two columns and the dependent variable is a dummy that takes a value of 1 if a violent event is observed in district i in the period [1979-1994]. OLS estimates are reported in the last two columns and the dependent variable is the (log) number of years with conflict episodes observed in district i in the period [1979-1994]. The variable *(# Schools/# Children)$_i$* corresponds to the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Further, given that in our main specifications we have a limited dependent variable, it is useful to check the sensitivity of our results when we abandon the underlying assumptions of the linear probability model and estimate a conditional logit specification instead. This is what we do in Online Appendix B.5. It turns out that we continue to find also for the conditional logit estimations a statistically significant conflict-reducing impact of the number of INPRES schools built.

Another alternative econometric specification that we implement is to move to a higher level of aggregation, at which some key political decisions may be taken. In particular, we collapse the data to a panel at the province-year level. As shown in Online Appendix B.6, even for this more coarse level of aggregation we find a strong and robust conflict-reducing impact of school construction.

Moreover, to address potential concerns that having *three* subperiods in part of the analysis may be somewhat ad hoc, we also display in Online Appendix B.7 the replication of the results of columns 7-9 of our baseline Table 2 when slicing the sample period in two, three, four or five subperiods. We continue to find school construction to have an increasing effect over time.

Finally, as a further sensitivity test in Online Appendix B.8 we also exploit the annual school construction levels, exploring the variation across districts between building more schools at the beginning versus at the end of the INPRES period. We find that our results remain very similar when we account for this further source of variation.

6.7 Robustness to using alternative conflict measures

As a first step to investigating the robustness to other conflict measures, we use the same data as in the main analysis, but focus in Appendix A.6 on the intensive instead of extensive margin of violence. While school construction turns out to have a strong effect on the extensive margin of whether or not conflict emerges, we also document that the INPRES program has reduced the intensity (i.e. frequency) of conflict episodes, as measured by the number of days, weeks or months with coverage of conflict. Similarly, when using the average length of newspaper articles covering an event as proxy for the incident's importance (intensity), we again find that schooling leads to not only fewer conflict incidents but also to more minor events (triggering shorter articles).

As a next step, we investigate robustness with respect to the set of keywords used. First of all, we carry out in Appendix A.7 a Monte Carlo analysis performing 1000 draws where we drop each time a third of our keywords. It turns out that even in this very demanding sensitivity test that (mechanically) drives down the number of conflict events detected, we continue to find a robust conflict-reducing effect of school construction.

Further, to investigate concerns about our findings having been obtained "by chance", we carry out in Online Appendix B.9 a placebo exercise where we randomly assign treatment in 1000 placebo datasets with the same average conflict likelihood as the "true" data. The results of this placebo exercise highlight how extremely unlikely it would have been to obtain our results "by chance".

We then study in Appendix A.8 the impact of extending (rather than narrowing down) the keywords used as well as relying on an alternative newspaper source, the *Canberra Times*, for constructing our conflict measure. Using the extended list of keywords (displayed in Appendix

A.8) may potentially reduce the risk of missing out on some conflict event but may well substantially increase the number of "false positives". Similarly, *Canberra Times* has a series of downsides with respect to the *Sydney Morning Herald*, as explained in Online Appendix B.1. Still it is important to assess the robustness with respect to these dimensions, and it is reassuring that in Appendix A.8 we find very similar results for the broadened keywords and the alternative newspaper source.

Further, we also carry out in Appendix A.9 a robustness analysis with respect to existing data from three datasets, GDELT (GDELT (2018)), ICEWS (ICEWS (2018)) and NVMS (NVMS (2019)). All of these datasets only cover time periods after INPRES school construction, which rules out any difference-in-difference analysis (and which is the reason why we had to collect and build our own conflict data in the first place). Still, the data allows us in Appendix A.9 to first show the high correspondence of our measure with the existing conflict data for the years of overlap, and then to replicate the analysis of the effect of school construction increasing over time (see e.g. the columns 4-6 of the baseline Table 2). When doing so in the Tables A13, A14 and A15 we reassuringly find similar results as in the baseline regressions.

6.8 Robustness with respect to geolocation

We also carry out additional sensitivity exercises with respect to the construction of our conflict measure. In particular, the results for alternative reclink scores adopted in the geographical matching of locations are reported in Online Appendix B.10, while Online Appendix B.11 implements an alternative mechanism for retrieving location information from newspaper reports. Reassuringly, in all cases our results are robust to these alternative ways of data construction.

6.9 Robustness to outliers and sample composition

Finally, in Online Appendix B.12 we investigate whether our results are driven by observations from a particular province or by a particular time period. We display graphically the evolution of fighting events by province over time and perform a regression analysis where we drop one-by-one observations from all 26 provinces in the sample, as well as modify the sample duration. Our results are found to be robust to these sensitivity exercises.

7 Heterogeneous Effects

Before analyzing the potential mechanisms at work, we shall in the current section present a series of findings with respect to heterogenous effects of our main estimates. First of all, it is interesting to see what kinds of events may be driving our results. For investigating this, we distinguish between keywords referring to disputes about "economic", "ethno-religious" or "political" dimensions.[17] As displayed in Table 4 below, our main findings are similar when focusing on any of theses three dimensions. These results could indicate that education matters throughout a wide range of different dimensions of conflict. They should however be interpreted with caution, as of course the absence of evidence of any heterogenous effects does not imply necessarily evidence of absence of any differences, as it may also be that our keyword distinctions are to coarse to pick up differential effects.

Table 4: Heterogeneous effects: Type of conflict events

	All Conflict		Economic		Ethnic-Religious		Political	
Dep. Variable: Conflict Episode$_{it}$ (Type)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0173***		-0.00780**		-0.00831**		-0.0118***	
	(0.00610)		(0.00353)		(0.00389)		(0.00430)	
*(# Schools / # Children)$_i$ * Years since 1978$_t$*		-0.00305***		-0.00182***		-0.00219***		-0.00257***
		(0.000705)		(0.000488)		(0.000572)		(0.000565)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.506	0.506	0.456	0.456	0.487	0.488	0.493	0.494
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.077	.077	.047	.047	.056	.056	.064	.064

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable in columns 1-2 is a dummy that takes a value of 1 if a violent event was observed in district i and year t. Estimates presented in columns 3 and 4 have as dependent variable a dummy of *economic* conflict events only, while columns 5 and 6 have as dependent variable a dummy of *ethno-religious* conflict events, and columns 7-8 *political* conflict events. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Another dimension of possible heterogeneity is whether results are different between districts previously not exposed to political violence and other districts where conflicts have occurred already in the period before the INPRES school construction. While in the former subset of districts conflict events could be seen as a new onset of fighting, in the latter subset of district any eruptions of violence may indicate the continuation of hostilities. The results of this sample split are shown in Table 5. Interestingly, we find that school construction under INPRES decreases the conflict likelihood both in districts with and without previous turmoil.

In the Online Appendix B.14 we study further heterogeneous effects, finding that our results hold both in rural and urban areas, as well as both in the presence and in the absence of the

[17] The keywords used for constructing the "economic", "ethno-religious", and "political" conflict variables are listed in Online Appendix B.13.

practice of bride prices.

Table 5: Heterogeneous effects: Conflict onset

	Districts with Conflicts pre-1979				Districts w/o Conflicts pre-1979			
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0230**	-0.0292*			-0.00722**	-0.00850**		
	(0.00901)	(0.0163)			(0.00300)	(0.00338)		
*(# Schools / # Children)$_i$ * Years since 1978$_t$*			-0.00238***	-0.00316**			-0.000771**	-0.00224***
			(0.000804)	(0.00153)			(0.000361)	(0.000819)
Observations	4,840	4,840	4,840	4,840	6,720	6,720	6,720	6,720
R-squared	0.309	0.490	0.309	0.490	0.095	0.342	0.096	0.345
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	No	Yes	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes	No	Yes
Sample Mean	.172	.172	.172	.172	.009	.009	.009	.009

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. Estimates presented in the first (resp., last) four columns are obtained restricting the sample to all districts where conflicts have (resp., have not) occurred in the period prior to the INPRES school construction. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

8 Channels and Mechanisms

After having studied above the impact of education on conflict, in the current section we shall investigate potential channels and mechanisms accounting for the conflict-reducing impact of education. The interest in understanding channels of transmission goes beyond academic curiosity, as knowing *how* and *why* education matters may unveil important implications for policy. For example, if we were to find any beneficial effects of education confined to act through economic incentives maybe it would be possible to achieve similar pacifying effects more cheaply, e.g. by promoting on the job training instead of school education. In contrast, if the main impact of education were to be related to fostering trust and understanding, again one could consider alternative policies delivering the same benefits. Finally, a conclusion of education affecting the scope for violence both through economic as well as societal channels of transmission may a priori make it more complicated to find a set of alternative policies achieving similar results.

8.1 Economic returns versus religious cleavages

We shall first carry out a "big-picture" comparison of potential economic versus societal forces at work. To do so, in Table 6 we start in column 1 by replicating the baseline specification of column 3 of the main Table 2, but interacting our usual variable of school construction with two variables that proxy for economic and societal channels of transmission.

In particular, we add the interaction of our explanatory variable of school construction with the level of religious polarization at the district level.[18] We expect a greater scope for education to matter through increased religious tolerance in districts where there is indeed substantial religious polarization. Put differently, if education were to deploy effects mostly in religiously homogenous places this would be harder to reconcile with mechanisms linked to religious tolerance than if the lion's share of education impact takes place in highly polarized places.

The second interaction term that we include is with a measure of returns to schooling. As we cannot draw on reliable pre-treatment wage data, we need to rely on proxies for wealth/income from survey data. Concretely, using data from the Indonesian population census of 1971 (IPUMS (2018)), we compute the relative economic advantages from having completed primary school (which is the case for roughly 45 percent of our sample) at the district level.[19] We focus first on the likelihood of living in a brick house (which is a superior housing quality capturing economic success), drawing on the answer to the survey question "*Dari apakah dinding luar dibuat*? (eng: Exterior wall material). In particular, our district-level returns to education measure corresponds to the formula

$$RoE[Bricks] = -\frac{\frac{Bricks(NoPS)}{Bricks(NoPS)+NoBricks(NoPS)}}{\frac{Bricks(PS)}{Bricks(PS)+NoBricks(PS)}},$$

where *Bricks(NoPS)* is the number of respondents in a district with brick housing and no completed primary school, and *NoBricks(NoPS)* is the number in a non-brick house without completed primary school. The definitions of *Bricks(PS)* and *NoBricks(PS)* are analogous, but simply for completed instead of non-completed primary school. Note that one advantage of this particular functional form is that the value of the measure is well-defined even when the value of *Bricks(NoPS)* is zero, i.e. when in a given district nobody without completed primary school lives in a brick house. In districts where this negative number is closer to zero, education has greater economic benefits in terms of housing quality. This variable is informative about economic mechanisms of transmission. If e.g. we systematically observe that education has the greatest pacifying effects in districts where it yields large economic benefits, it is more likely that economic mechanisms (such as education increasing the opportunity cost of fighting) are at work than if economic returns are unrelated to the pacifying force of education.

[18] We build this variable using the data from the Indonesian population census of 1971 (IPUMS (2018)), and applying the widely used polarization measure described in Montalvo and Reynal-Querol (2005).

[19] In our sample we focus on all adults, defined as individuals of at least 21 years of age, which corresponds to the 95th percentile of the age distribution of individuals still attending school. Put differently, this sample restriction allows us to have almost exclusively individuals in the sample who are not currently enrolled in schooling. If we use the full sample or we drop all individuals attending school our main results are similar. Corresponding tables are reported in Online Appendix B.15.

As shown in column 1 of Table 6, both our proxies for economic and societal mechanisms have the expected negative signs and are statistically significant. In particular, we find that schooling in general decreases the scope for conflict, but this effect gets magnified in religiously polarized areas and in places with large economic returns to education. Put differently, while education curbs conflict everywhere, it particularly does so in religiously diverse districts, as well as in places where living standards are relatively sensitive to the level of schooling achieved. In the goal of investigating the sensitivity of the returns to education proxy used, column 2 carries out a similar exercise as column 1, but replaces the returns to education variable based on brick housing with another measure of educational returns based on entrepreneurship. The variable construction is analogous, simply the question used is not the one mentioned above but instead "*Bekerja sebagai apa*?" (eng: Occupation Status). The results are similar, but now the returns to education proxy is not statistically significant.

In column 3 we use as other alternative returns to education proxy the simple average of the bricks and entrepreneurship measures used in the columns 1 and 2, respectively. Again, the interaction term with religious polarization keeps being negative and robustly significant, while the returns of education proxy has the expected negative sign but narrowly misses statistical significance. Similarly, in column 4 we make use of principal component analysis (PCA) to creating a joint measure of education returns englobing both the bricks and entrepreneurship data. Again, the interaction term of this alternative variable has the expected negative sign but misses the statistical significance threshold.

Columns 5 to 8 perform analogous regressions, but focusing on the second main specification of our baseline specifications. We replicate the baseline specification of column 6 of the main Table 2, but interacting our usual variable of school construction with the aforementioned proxies for economic and societal channels of transmission. This specification allows to understand up to what extent the effect of INPRES school construction is increasing over time, and the interaction terms enable us to perceive the dynamic evolution of economic and societal mechanisms. We find that not only religious polarization, but also the various proxies for economic returns to schooling increase substantially and robustly the long-run conflict-reducing pattern of education.[20]

[20]Using the heterogeneous effects coefficients of column 7 of Table 6 and the district specific measures of returns to education and religious polarization, we find that in 92 percent of districts (i.e. 240 out of 262) education has overall the expected negative (conflict-reducing) sign. Interestingly, in the remaining 22 districts the lack of conflict-reducing effect is driven by their very low economic returns to education (i.e. when attributing to them the sample average returns to education, for all of them their district-specific coefficient of education would also turn to a negative (conflict-reducing) sign). This result that very low economic opportunities can jeopardize the pacifying effect of education is in line with the recent results of Campante and Chor (2012) and Campante and Chor (2014).

Table 6: Mechanism: Economic and societal channels

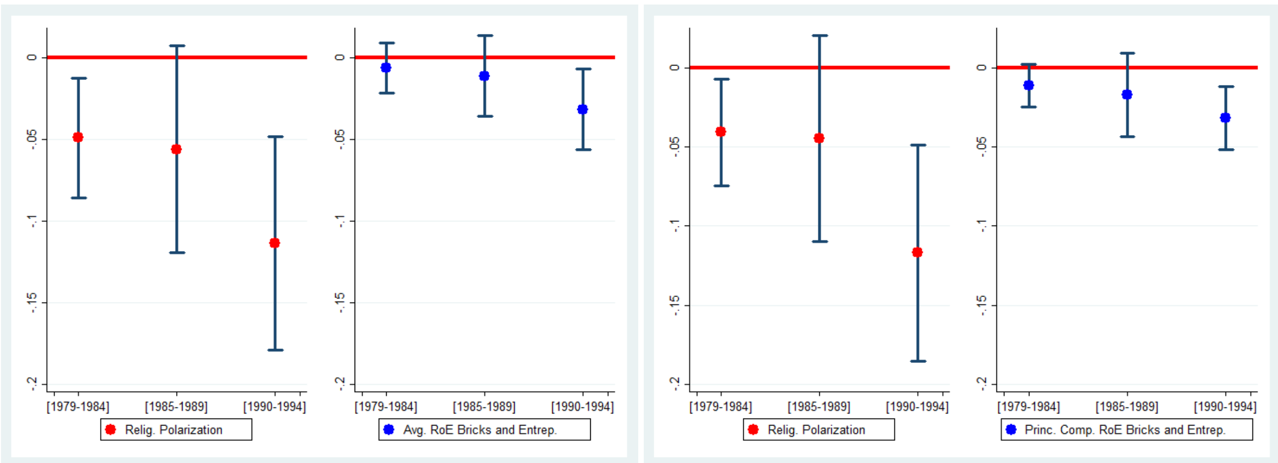
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0133**	-0.00651	-0.00996	-0.00897				
	(0.00661)	(0.00585)	(0.00621)	(0.00588)				
*(# Schools / # Children)$_i$ * Years since 1978$_t$*					-0.00312***	-0.00268***	-0.00296***	-0.00286***
					(0.000909)	(0.000720)	(0.000777)	(0.000831)
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Religious Polarization$_i$*	-0.0392*	-0.0351*	-0.0473**	-0.0375*				
	(0.0214)	(0.0192)	(0.0211)	(0.0203)				
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Return to Education [Bricks]$_i$*	-0.0236*							
	(0.0121)							
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Return to Education [Entrep.]$_i$*		-0.00111						
		(0.00474)						
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Average RoE Bricks and Entrep.$_i$*			-0.00633					
			(0.00868)					
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*				-0.0116				
				(0.00821)				
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Religious Polarization$_i$*					-0.00767***	-0.00711***	-0.00799***	-0.00845***
					(0.00277)	(0.00233)	(0.00231)	(0.00257)
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Return to Education [Bricks]$_i$*					-0.00347**			
					(0.00147)			
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Return to Education [Entrep.]$_i$*						-0.00103**		
						(0.000494)		
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Average RoE Bricks and Entrep.$_i$*							-0.00224**	
							(0.000921)	
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*								-0.00233***
								(0.000850)
Observations	9,480	9,040	10,480	8,040	9,480	9,040	10,480	8,040
R-squared	0.522	0.529	0.517	0.534	0.523	0.530	0.519	0.536
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.081	.076	.079	.078	.081	.076	.079	.078

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Religious Polarization* corresponds to the level of religious polarization in district i, whereas the variable *Return to Education* (*RoE*) indicates the relative economic advantages at the district level from having completed primary school. Religious polarization and returns to education measures were computed using the 1971 Census (IPUMS (2018)) (see additional details in the text). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To compare the size of the coefficients of the interaction terms with polarization versus with economic returns to education, we can focus on the estimates of our preferred column 7 of Table 6. Consider the differential effect on the conflict probability after 10 years of one additional school in a district with 75th percentile versus median religious polarization. One more school reduces the conflict risk in the more polarized district by around 3 percentage points. Now we perform the analogous exercise for the differential effect of being in a district with returns to education at the 75th percentile versus at the median, which yields an additional 0.6 percentage point drop in the conflict risk for one more school in the district with higher returns to schooling. Thus, while both higher religious polarization and greater economic returns to education magnify our baseline effect, we find a quantitatively larger impact for the former than for the latter.[21]

Taken together, this means that while higher religious polarization immediately magnifies any effects of education reducing fighting, the importance of economic returns takes longer to kick in. In the very short-run, the modulating effect of economic returns to education is not very big and only borderline statistically significant, while in the long-run economic forces have a robust impact on boosting the conflict-reducing effect of education. This novel result is further substantiated in Figure 3 below which displays the evolution of the coefficients of interest when slicing our sample in three subperiods. While in the first two subperiods economic returns to education do not matter greatly, in the last subperiod they become a more important modulating factor of the impact of school construction on conflict.

Figure 3: Impact of economic returns and religious cleavages over time



A) Avg. RoE Bricks and Entrep. B) Princ. Comp. RoE Bricks and Entrep.

Note: The figures plot the evolution of the coefficients *(# Schools / # Children)*$_i$ * *Religious Polarization*$_i$ and *(# Schools / # Children)*$_i$ * *Return to Education*$_i$ over three subperiods [(1979-1984), (1985-1989) and (1990-1994)]. Estimates reported in Panel A and Panel B are those obtained using the specifications displayed in columns 3 and 4 of Table B22 in Online Appendix B.15, respectively. Estimates reported in the right figure of Panel A (Panel B) are those obtained using as measure of returns to education the simple average (PCA) of bricks and entrepreneurship measures. Religious polarization and returns to education measures were computed using the 1971 Census (IPUMS (2018)) (see additional details in the text).

[21] The effect for religious polarization is computed as -.00799 * 10 * 0.343 = -.027, while for returns to education the effect is computed as -.00224 * 10 * 0.260 = -.0057.

In further robustness and sensitivity exercises, reported in Online Appendix B.15, we assess whether the aforementioned results on channels and mechanisms are robust to alternative ways of constructing the conflict variable, finding that the results remain very similar.

Overall, we take the findings of Table 6 and Online Appendix B.15 as evidence that both economic as well as societal channels of transmission may be at work when it comes to linking schooling to a reduction in fighting. While the impact of education on curbing conflict in religiously polarized places is immediate, the importance of economic returns to schooling takes some years to invigorate the effects of school construction.

8.2 How education may attenuate religious tensions

In the last subsection we have found that education may well work through both economic channels of transmission (i.e. a higher opportunity cost of conflict), as well as through societal mechanisms, i.e. by making religious polarization matter less. In the current subsection we shall now investigate in greater depth how education may be able to reduce the scope for religious polarization to fuel fighting.

In particular, we make use of the wave 5 of the IFLS Survey, conducted in 2014 in 228 districts, to investigate the effects of the INPRES program on both i) religious tolerance and ii) community participation.[22] In doing so, we focus on answers provided by individuals born between 1945 and 1972. Our identification strategy relies on the fact that the date of birth and the region of birth jointly determine exposure to the school construction program. All children born in 1962 or before did not benefit from the program since they were too old to enroll in newly constructed schools when the program started.[23]

As main dependent variables we use the following three survey questions:

- Religious Tolerance I - **Trust**: Question: *Taking into account the diversity of religions in the village, I trust people with the same religion as mine more.* [0-Strongly agree, 1-Agree, 2-Disagree, 3-Strongly disagree]

[22] Our analysis in this subsection is related to Roth and Sumarto (2015), who study –using alternative survey data– the impact of school construction on the answer to the question "What is your opinion on an activity done in your neighborhood by a group of people which are from a different ethnicity, resp. religion". In line with our results, they conclude that schooling fosters inter-group tolerance. However, we find their results hard to interpret, given how vague the underlying survey question is formulated, as a person's opinion on an unspecified "activity" may well depend on the nature of the actual activity, with e.g. economic competition being different from social interaction.

[23] Note that this individual-level identification strategy follows the one adopted by Duflo (2001). The only difference is that Duflo (2001) included the cohorts 1950-1973, while we extend the pre-treatment sample by five years in order to balance the number of treated and control individuals. Crucially, our results are very similar if we restrict like Duflo (2001) the sample to post-1950 individuals, as shown in Online Appendix B.16.

- Religious Tolerance II - **Marriage**: Question: *How do you feel if someone with different faith from you marry one of your close relatives or children?* [0-Strongly objected, 1-Objected, 2-No Objection, 3-No Objection at all]

- Community Participation - **Arisan**: *Have you participated in Arisan in the last 12 months?* [0 - No, 1 - Yes]

Note that "Arisan" is an Indonesian form of a rotating savings and credit association (roscas) (see e.g. the discussion in Miguel et al. (2006)). While participation to roscas is partly driven by purely economic forces, such community credit groups have typically still be seen as associated with social capital and strong local ties (see Putnam et al. (1993), Miguel et al. (2006), Anderson et al. (2009)).

The aforementioned Trust and Marriage variables are used as continuous variables ranging from 0 to 3, treating the scales of the survey questions as cardinal. We also code a dichotomous version of these variables (TrustD and MarriageD), with values 0-1 being coded as 0, and values 2-3 coded as 1). The results for these dichotomous measures are very similar to the main results, and are relegated to Online Appendix B.16. The "Arisan" (roscas) variable is a dummy (0-1).

We estimate the effect of the INPRES program on the above-mentioned variables ($Survey_n$) using the econometric specification

$$Survey_n = \alpha + \beta \frac{\#Schools\ Built}{\#Children}_i * Born\ after\ 1962_{nic} + FE_i + FE_c + \epsilon_n,$$

where $(\#Schools\ Built/\#Children)_i$ represents the number of primary schools constructed under the INPRES program per 1,000 children in district i, the variable *Born after 1962* is a dummy that takes value 1 if the individual n of cohort c was born after 1962 in district i, and the vectors of FE_i and FE_c represent district and cohort FEs, respectively.

Table 7 below displays the results. We find that having been exposed to more intensive INPRES school construction (i.e. being born in a place with more INPRES schools built *and* being part of a birth cohort affected by it) results in an increase in all three aforementioned indicators. Put differently, we find that being exposed to more intensive school construction boosts religious tolerance, as measured both in terms of trust (column 1) as well as marriage approval (column 2) with respect to people of different faith. Similarly, we also find that more intensive school construction exposure leads to a greater propensity to participate to "Arisan" community credit groups (roscas) (column 3). These results prove robust when replicated in columns 4-6 with gender, ethnicity and religion fixed effects. Overall, the findings of Table 7 highlight that indeed education may be able to reduce religious intolerance and foster local community interaction.

These results are robust to a series of sensitivity tests relegated to Online Appendix B.16, where we vary the coding of the survey answers, where we modify the measurement of religiosity and where we modify the starting date of the cohort window applied.

Crucially, in Online Appendix B.16 we also investigate whether this tolerance-boosting effect of education is confined to particular religions. We find that this is not the case, and that for all religions in the sample (Islam, Christianism, Others) we find that education contributes to increasing inter-religious trust and tolerance.

Table 7: Societal channels: Religious tolerance and local community involvement

Dep. Variable:	(1) $Trust_n$	(2) $Marriage_n$	(3) $Roscas_n$	(4) $Trust_n$	(5) $Marriage_n$	(6) $Roscas_n$
*(# Schools / # Children)$_i$ * Born after 1962$_n$*	0.0344**	0.0321**	0.0189***	0.0322**	0.0263**	0.0166***
	(0.0140)	(0.0127)	(0.00665)	(0.0137)	(0.0114)	(0.00607)
Observations	10,521	10,522	11,229	10,521	10,522	10,461
R-squared	0.107	0.179	0.154	0.134	0.223	0.237
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes	Yes
Gender FEs	No	No	No	Yes	Yes	Yes
Ethnicity FEs	No	No	No	Yes	Yes	Yes
Religion FEs	No	No	No	Yes	Yes	Yes

NOTE: The unit of observation is an individual n born in district i. The sample covers all individuals surveyed in the *Wave 5* of the IFLS SURVEY, born between 1945 and 1972. OLS estimates are reported in all columns. $Trust_n$ and $Marriage_n$ variables are used as continuous variables ranging from 0 to 3, treating the scales of the survey questions as cardinal. $Roscas_n$ is a dummy that take a value of 1 if the individual participated to a *arisan* community group over the previous 12 months. Additional details on survey variables are provided in the text. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Born after 1962$_n$* is a dummy that takes a value of 1 if a given individual n was born after 1962 in district i. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

An important remaining question is whether the reason education manages to attenuate religious tensions is due to e.g. inculcating values of tolerance or, more trivially, that education may simply make religion matter less. Put differently, if educated people become less religious in general (as found e.g. for Canada by Hungerman (2014)), there is (mechanically) less scope for religious conflict. Table 8 below allows us to answer this question. In particular, we explore in column 1 whether the program affected the level of religiosity of the individuals surveyed in the wave 5 of the IFLS Survey, drawing on the question: "How religious are you" [3 - Very Religious, 2 - Somewhat religious, 1 - Rather religious, 0 - Not religious]. We again focus in the main analysis on the continuous scale, but show robustness in Online Appendix B.16 to a dichotomous version of the religiosity measure. As a second measure, in column 2 we also draw on the question “How many times do you pray each day?" [coded as 1 if the answer was "Given times", and as 0 for "Not every day" "Do not practice"]. The columns 1-2 are replicated in columns 3-4, but including also gender, ethnicity and religion fixed effects.

Interestingly, the results of Table 8 overall indicate that religious beliefs are not affected by school construction, which may be interpreted as evidence that the education-induced decrease in religious intolerance is not purely due to educated people losing their faith, but could possibly be driven by a genuine increase in tolerance. These findings could be related to the fact that education content in Indonesia contained some teaching of the principles of the state ideology *Pancasila* that stresses at the same the time the importance of religious faith as well as promotes religious tolerance (see Nishimura (1995)).

Table 8: School construction and religiosity

	(1)	(2)	(3)	(4)
Dep. Variable:	$Religiosity_n$	$Prayers_n$	$Religiosity_n$	$Prayers_n$
*(# Schools / # Children)$_i$ * Born after 1962$_n$*	0.0206*	-0.00547	0.0180	-0.00531
	(0.0120)	(0.00474)	(0.0130)	(0.00478)
Observations	10,495	9,292	10,495	9,292
R-squared	0.093	0.074	0.104	0.088
District FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Gender FEs	No	No	Yes	Yes
Ethnicity FEs	No	No	Yes	Yes
Religion FEs	No	No	Yes	Yes

NOTE: The unit of observation is an individual n born in district i. The sample covers all individuals surveyed in the *Wave 5* of the IFLS SURVEY, born between 1945 and 1972. OLS estimates are reported in all columns. $Religiosity_n$ is a continuous variable ranging from 0 to 3, treating the scale of the survey question as cardinal. $Prayers_n$ is a dummy that takes a value of 1 if the individual prays every day. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Born after 1962$_n$* is a dummy that takes a value of 1 if a given individual n was born after 1962 in district i. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the Online Appendix B.17 we study the impact of schooling on relations between ethnic instead of religious groups. We detect no effect of schooling on ethnic tolerance, which is in line with the particular emphasis of the state ideology *Pancasila* on religious tolerance and freedom, which is not analogously present for inter-ethnic relations.

8.3 Voice versus violence

As found above, education tends to boost civic involvement. Hence a natural question to ask is if education may make people express grievances and discontent in different ways, providing educated citizens with both the incentives and means for voicing discontent in a peaceful manner rather than resorting to violence. As discussed above, the existing evidence linking education to the propensity for participating to protests is mixed, with Campante and Chor (2012) and Campante and Chor (2014) finding that education boosts the willingness to protest, while Passarelli and Tabellini (2017) find the opposite effect. In what follows, we do not only restrict our focus to protests per se, but study the *relative* predominance of peaceful versus violent forms of contesting authority.

In particular, we start by creating a measure of peaceful protest from newspaper articles of the Sydney Morning Herald, proceeding analogously as for the construction of our conflict measure, but making use of a different set of keywords.[24] Reassuringly, we find a very large overlap with the protest measure of GDELT for the years when both variables are available: In 95 percent of cases, both measures agree on the coding of a given district-year of having or not having a peaceful protest. All details of the data construction and analysis and additional results are presented in Online Appendix B.18.

Note that besides studying the interesting trade-off between voice and violence this analysis also serves the purpose of a "placebo"-type robustness check: Imagine that for some reason (mechanically) newspapers were to cover after 1979 differently places with higher INPRES school construction (remember that general time invariant differences in reporting are controlled for by the district fixed effects). Such hypothetical biases from news coverage could to a similar extent also apply to other events such as peaceful protests and we would expect similar results as for violent events. In contrast, finding different results for peaceful and violent activities would a priori be reassuring, and make this type of mechanical reporting bias less likely.

As shown in Table 9, it turns out that school construction reduces violent conflict events (columns 1-2), but only has a small, not statistically significant effect on peaceful protests and demonstrations (columns 3-4). Taking "voice" and "violence" together in the same specification, by coding as dependent variable the difference between peaceful and violent events, we find that education statistically significantly increases this wedge between peaceful and violent means of opposition to the state (columns 5-6). Thus, in a nutshell, we indeed find that schooling tends to move resistance from violent to more peaceful modes of expression. This result is confirmed by the supplementary analysis performed in Online Appendix B.18 where we replicate the analysis for alternative datasets. We always find that the impact of schooling on peaceful protests is quantitatively small and sometimes non-significant, and that the relative importance of conflict with respect to peaceful protests goes down in the aftermath of more INPRES school construction. This is again consistent with the conclusion that education pushes modes of opposition from violence towards voice.

9 Conclusions

This paper is to the best of our knowledge the first one to study the causal impact of education on conflict. We have exploited in a difference-in-difference specification the variation in school construction under the INPRES program from 1974 to 1978 in Indonesia. In order to be able to

[24]In particular, we used the following keywords: "protest", "demonstration", "march", "gather", "manifestation", "picket".

Table 9: Conflict events vs pacific events

Panel A

	Conflict Events		Pacific Events		Δ *Pacific - Conflict*	
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0127***	-0.00173***	-5.73e-05	-0.00202	0.0127***	0.0153*
	(0.00448)	(0.00610)	(0.00306)	(0.00641)	(0.00389)	(0.00887)
Observations	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.506	0.279	0.405	0.183	0.325
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes
Sample Mean	.08	.08	.02	.02	-.06	-.06

Panel B

	Conflict Events		Pacific Events		Δ *Pacific - Conflict*	
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
*(# Schools / # Children)$_i$ * Years since 1978$_t$*	-0.00146***	-0.00305***	-6.54e-05	4.39e-05	0.00140***	0.00309***
	(0.000421)	(0.000705)	(0.000230)	(0.000430)	(0.000347)	(0.000762)
Observations	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.346	0.506	0.279	0.405	0.184	0.326
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes
Sample Mean	.08	.08	.02	.02	-.06	-.06

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. In columns 1 and 2 the dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. Columns 3 and 4 have as dependent variable a dummy that takes a value of 1 if a peaceful protest was observed in district i and year t. In columns 5 and 6 the dependent variable is the difference between peaceful and violent events observed in district i and year t. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict and protest data was constructed using the *Sydney Morning Herald*, following the approach described in Sections 4.1 and 8.3. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

analyse its effects on conflict, we had to collect our own dataset on conflict at the level of the 289 districts in Indonesia over the period 1955-1994. Applying up-to-date webcrawling, scraping and text recognition to information from over 820,000 newspaper pages, we have built a novel and very extensive data set on political violence in Indonesia.

We have found that school construction in Indonesia has indeed had a statistically significant and quantitatively substantial conflict-reducing impact which survives extensive robustness checks with respect to estimator, specification, measures, data construction and potential confounders. We detect that schooling matters both for areas with and without previous fighting and that economic, ethno-religious and political conflict is reduced alike.

In terms of the underlying mechanisms our results indicate that both larger religious polarization and greater economic returns to schooling magnify the beneficial effects of education, and that while societal mechanisms appear at work right away, economic factors start impacting mostly after some years. Studying individual survey data on inter-religious trust, local community involvement and religiosity, we find that education leads to higher trust and tolerance of other religious groups. We do not detect any impact on religiosity, ruling out that higher inter-religious trust and tolerance could be mechanically driven by a drop in religious beliefs. We also detect that schooling leads to a shift from violence to voice. Taken together, our findings suggest that education expansion may yield substantial benefits in terms of conflict prevention that go well beyond the narrow economic human capital gain of schooling.

We very much encourage further research on this topic. While the results of schooling expansion on reduced civil conflict in Indonesia are very telling, it would be important to analysis the impact of schooling in very different contexts. In particular, our context features the impact of primary school expansion with a curriculum focused on secular teaching of basic skills and promoting –if anything– religious tolerance. The benefits of school construction may be different for secondary or tertiary schooling or in settings where the curriculum promotes values of inter-group intolerance and defamation. Hence, an under-studied topic in the literature seems to be the impact of educational *content* on the scope for civil conflict. Further, as for development, the impact of education on interstate wars could be potentially different than for civil wars (while opportunity costs reduce in both cases incentives for fighting, in the case of interstate wars large conflicts necessitate advanced fighting capabilities which may be built up more easily in more developed and educated societies). Thus, also studying the impact of education on international wars seems an important gap in the literature.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132.

Aghion, P., Jaravel, X., Persson, T., and Rouzet, D. (2019). Education and military rivalry. *Forthcoming in the Journal of the European Economic Association.*

Akresh, R., Halim, D., and Kleemans, M. (2018). Long-term and intergenerational effects of education: Evidence from school construction in Indonesia. *NBER Working Paper Number 25265.*

Algan, Y., Cahuc, P., and Shleifer, A. (2013). Teaching practices and social capital. *American Economic Journal: Applied Economics*, 5(3):189–210.

Alsan, M. M. and Cutler, D. M. (2013). Girls' education and hiv risk: Evidence from Uganda. *Journal of Health Economics*, 32(5):863–872.

Anderson, S., Baland, J.-M., and Moene, K. O. (2009). Enforcement in informal saving groups. *Journal of Development Economics*, 90(1):14–23.

Ashraf, N., Bau, N., Nunn, N., and Voena, A. (2019). Bride price and female education. *Forthcoming in the Journal of Political Economy*.

Barakat, B. and Urdal, H. (2009). Breaking the waves? Does education mediate the relationship between youth bulges and political violence? *Working Paper*.

Bazzi, S., Blair, R., Blattman, C., Dube, O., Gudgeon, M., and Peck, R. (2019). The promise and pitfalls of conflict prediction: Evidence from indonesia and colombia. *Working Paper*.

Bazzi, S. and Gudgeon, M. (2018). The political boundaries of ethnic divisions. *NBER Working Paper Number 24625*.

BBC (2018). *Indonesia profile and timeline*. https://www.bbc.com/news/world-asia-pacific-15114517.

Behrman, J. A. (2015). The effect of increased primary schooling on adult women's hiv status in Malawi and Uganda: Universal Primary Education as a natural experiment. *Social Science & Medicine*, 127:108–115.

Berman, N., Couttenier, M., Rohner, D., and Thoenig, M. (2017). This mine is mine! How minerals fuel conflicts in Africa. *American Economic Review*, 107(6):1564–1610.

Berrebi, C. (2007). Evidence about the link between education, poverty and terrorism among Palestinians. *Peace Economics, Peace Science and Public Policy*, 13(1).

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Bharati, T., Chin, S., and Jung, D. (2017). Recovery from an early life shock through improved access to schools: Evidence from indonesia.

Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.

Breierova, L. and Duflo, E. (2004). The impact of education on fertility and child mortality: Do fathers really matter less than mothers? *NBER Working Paper Number 10513*.

Brown, C. (2003). *A Short History of Indonesia: The Unlikely Nation?* Allen & Unwin.

Campante, F. R. and Chor, D. (2012). Why was the Arab world poised for revolution? Schooling, economic opportunities, and the Arab Spring. *Journal of Economic Perspectives*, 26(2):167–88.

Campante, F. R. and Chor, D. (2014). "The people want the fall of the regime": Schooling, political protest, and the economy. *Journal of Comparative Economics*, 42(3):495–517.

Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2017). Curriculum and ideology. *Journal of Political Economy*, 125(2):338–392.

Cantoni, D. and Yuchtman, N. (2013). The political economy of educational content and development: Lessons from history. *Journal of Development Economics*, 104:233–244.

CIA (2018). *The World Factbook 2018*. Government Printing Office.

Clots-Figueras, I. and Masella, P. (2013). Education, language and identity. *The Economic Journal*, 123(570):F332–F357.

Collier, P. (2008). *The Bottom Billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press.

Collier, P. and Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford economic papers*, 56(4):563–595.

Company, B. P. (2018). *BP Statistical Review of World Energy*. British Petroleum Company.

Cribb, R. (2010). *Digital Atlas of Indonesian History*. Nordic Institute of Asian Studies.

De la Briere, B., Filmer, D., Ringold, D., Rohner, D., and Denisova, A. (2017). *From Mines and Wells to Well-Built Minds: Turning Sub-Saharan Africa's Natural Resource Wealth Into Human Capital*. World Bank Publications.

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*, 88(9-10):1697–1720.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4):795–813.

Duflo, E. (2004). The medium run effects of educational expansion: Evidence from a large school construction program in Indonesia. *Journal of Development Economics*, 74(1):163–197.

Duflo, E., Dupas, P., and Kremer, M. (2015). Education, hiv, and early fertility: Experimental evidence from Kenya. *American Economic Review*, 105(9):2757–97.

Encycl. Britannica (2018). *Encyclopaedia Britannica*.

Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1):75–90.

Friedman, W., Kremer, M., Miguel, E., and Thornton, R. (2016). Education as liberation? *Economica*, 83(329):1–30.

GDELT (2018). GDELT dataset. https://www.gdeltproject.org/.

Glaeser, E. L., Ponzetto, G. A., and Shleifer, A. (2007). Why does democracy need education? *Journal of Economic Growth*, 12(2):77–99.

Hainmueller, J. and Hiscox, M. J. (2007). Educated preferences: Explaining attitudes toward immigration in Europe. *International Organization*, 61(2):399–442.

Humphreys, M. and Weinstein, J. M. (2008). Who fights? The determinants of participation in civil war. *American Journal of Political Science*, 52(2):436–455.

Hungerman, D. M. (2014). The effect of education on religion: Evidence from compulsory schooling laws. *Journal of Economic Behavior & Organization*, 104:52–63.

ICEWS (2018). Integrated Crisis Early Warning System (ICEWS) dataset. https://dataverse.harvard.edu/dataverse/icews.

IPUMS (2018). Population census of the Central Bureau of Statistics of the Republic of Indonesia, 1971, 1980, and 1990. https://international.ipums.org/international/.

Kim, H. B., Choi, S., Kim, B., and Pop-Eleches, C. (2018). The role of education interventions in improving economic rationality. *Science*, 362(6410):83–86.

Krueger, A. B. and Malečková, J. (2003). Education, poverty and terrorism: Is there a causal connection? *Journal of Economic Perspectives*, 17(4):119–144.

Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, 94(1):155–189.

Martinez-Bravo, M. (2017). The local political economy effects of school construction in Indonesia. *American Economic Journal: Applied Economics*, 9(2):256–89.

Merlino, L. P., Steinhardt, M. F., and Wren-Lewis, L. (2019). More than just friends? school peers and adult interracial relationships. *Forthcoming in the Journal of Labor Economics.*

Miguel, E., Gertler, P., and Levine, D. I. (2006). Does industrialization build or destroy social networks? *Economic Development and Cultural Change*, 54(2):287–317.

Milligan, K., Moretti, E., and Oreopoulos, P. (2004). Does education improve citizenship?

Evidence from the United States and the United Kingdom. *Journal of Public Economics*, 88(9-10):1667–1695.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.

Morgan, R. (2018). Roy Morgan research newspaper readership numbers. http://www.roymorgan.com/industries/media/readership/newspaper-readership.

Nishimura, S. (1995). The development of Pancasila moral education in Indonesia. *Japanese Journal of Southeast Asian Studies*, 33(3):303–316.

NVMS (2019). National Violence Monitoring System (NVMS) dataset. http://snpk.kemenkopmk.go.id.

Osili, U. O. and Long, B. T. (2008). Does female schooling reduce fertility? Evidence from Nigeria. *Journal of Development Economics*, 87(1):57–75.

Østby, G. and Urdal, H. (2011). Education and civil conflict: Areview of the quantitative, empirical literature. *Background paper prepared for the Education for All Global Monitoring Report*.

Passarelli, F. and Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, 125(3):903–946.

Postlethwaite, T. N. and Thomas, R. M. (2014). *Schooling in the ASEAN region: Primary and secondary education in Indonesia, Malaysia, the Philippines, Singapore, and Thailand*. Elsevier.

Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.

Refworld (2001). *Child Soldiers Global Report 2001 - Indonesia*. Child Soldiers International.

Rohner, D. (2016). Barrels, books, and bullets: How education can prevent conflict and promote development in resource rich countries. *mimeo, University of Lausanne*.

Rohner, D. (2017). The economics of conflict and peace. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*.

Roth, C. and Sumarto, S. (2015). Does education increase interethnic and interreligious tolerance? Evidence from a natural experiment. *Working Paper*.

Saia, A. (2017). Choosing the open sea: The cost to the UK of staying out of the euro. *Journal of International Economics*, 108:82–98.

Shemyakina, O. (2011). The effect of armed conflict on accumulation of schooling: Results from Tajikistan. *Journal of Development Economics*, 95(2):186–200.

Somanathan, A. (2008). *Use of modern medical care for pregnancy and childbirth care: Does female schooling matter?* The World Bank.

Strauss, J., Witoelar, F., and Sikoki, B. (2016). The fifth wave of the Indonesia family life survey: Overview and field report. *Working Paper*.

Tezcür, G. M. (2016). Ordinary people, extraordinary risks: Participation in an ethnic rebellion. *American Political Science Review*, 110(2):247–264.

Thyne, C. L. (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.

Tollefsen, A. F., Strand, H., and Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, 49(2):363–374.

Verwimp, P. and Van Bavel, J. (2013). Schooling, violent conflict, and gender in Burundi. *The World Bank Economic Review*, 28(2):384–411.

Wantchekon, L., Klašnja, M., and Novta, N. (2014). Education and human capital externalities: Evidence from colonial Benin. *The Quarterly Journal of Economics*, 130(2):703–757.

A Appendix

A.1 Empirical Strategy and Main Results: Balancing Covariates

As discussed above in section 5.2 we study in what follows the potential determinants of the intensity of school construction under INPRES. The official rule for INPRES school construction corresponded in building more schools in places with initially fewer schools to equalize the school density across different regions in Indonesia (see Duflo (2001)). We observe that indeed to a substantial extent this rule was followed: Pre-enrollement rates and new school construction correlates at -0.13 (significant at the 1% level). The rest of the variation is likely due to measurement error, imperfect implementation and/or random factors.

To investigate this further, we carry out a regression analysis of determinants of INPRES school construction. As shown in Table A1, the only correlate of school construction that is found to strongly determine the numbers of schools built is the pre-INPRES enrollment rate of school-aged children. As displayed in the Appendix A.2 below, the results of this paper are robust to controlling for pre-INPRES school enrollment and the other socio-economic correlates (with, as expected, a slightly lower coefficient value).

Table A1: Balancing covariates - Dep. variable: *(log) # INPRES schools*

Dep. Variable: (log) INPRES Schools$_i$	(1)	(2)	(3)	(4)	(5)	(6)
(log) Children 5-14$_i$	0.731***	0.731***	0.735***	0.736***	0.736***	0.736***
	(0.0267)	(0.0268)	(0.0270)	(0.0272)	(0.0271)	(0.0272)
(log) School Attendance$_i$	-1.042***	-1.033***	-1.011***	-1.020***	-0.999***	-1.008***
	(0.261)	(0.262)	(0.262)	(0.262)	(0.264)	(0.263)
(log) Enrollment Population$_i$	0.0427	0.0413	0.0455	0.0491	0.0437	0.0475
	(0.0689)	(0.0691)	(0.0689)	(0.0692)	(0.0691)	(0.0694)
(log) Rural Population$_i$	0.0801	0.0688	0.0834	0.0808	0.0700	0.0677
	(0.171)	(0.173)	(0.171)	(0.171)	(0.173)	(0.173)
(log) Primary Industries Employment$_i$	0.0976		0.103	0.105		
	(0.197)		(0.197)	(0.197)		
(log) Mining Employment$_i$		-0.267			-0.359	-0.341
		(0.955)			(0.958)	(0.958)
(log) Agricultural Employment$_i$		0.111			0.119	0.120
		(0.199)			(0.199)	(0.199)
Dummy Conflict [Pre-1979]$_i$			-0.0509		-0.0528	
			(0.0461)		(0.0463)	
(log) Years with Conflict [Pre-1979]$_i$				-0.0282		-0.0292
				(0.0283)		(0.0284)
Observations	289	289	289	289	289	289
R-squared	0.802	0.802	0.803	0.803	0.803	0.803

NOTE: The unit of observation is a district i. The sample covers 289 districts. OLS estimates are reported in all columns. The dependent variable is the (log) number of primary schools constructed under the INPRES program in a district i. The variable *(log) Children 5-14$_i$* represents the number of school-aged children in district i. The variable *(log) Enrollment Population$_i$* represents the population-wise pre-INPRES enrollment rates observed in district i. The variable *(log) Rural Population$_i$* represents the share of population of district i living in rural areas. The variable *(log) Primary Industries Employment$_i$* represents share of population working in primary industries (i.e. Agricultural and Mining Industries) observed in district i. All socio-economic variables were computed using the 1971 Census (IPUMS (2018)). The variable *Dummy Conflict [Pre-1979]$_i$* is a dummy that takes a value of 1 if a violent event is observed in district i in the period [1955-1979]. The variable *(log) Years with Conflict [Pre-1979]$_i$* is the (log) number of years with conflict episodes observed in district i in the period [1955-1979]. Standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2 Robustness Analysis: Controlling for Major Socio-Economic Covariates

As discussed in detail in section 6.2, we control below in the Table A2 for the interactions of the "post-1978" and "years since 1978" variables with a wide array of socio-economic controls.[25] Reassuringly, the results are robust (yet, as expected, with a slightly lower coefficient size), no matter which additional controls we add.[26]

[25] In the Table A2 these socio-economic controls are constructed from census data (IPUMS (2018)). Note that our results are very similar and remain statistically significant at the 1 percent level if we rely instead on analogous data provided by Duflo (2001), leading however to more missing observations (results available upon request).

[26] Another approach would be to run the "reduced form" regression estimating directly the impact on conflict of pre-INPRES school enrollment (which is a major determinant of INPRES school construction that in turn affects conflict). As discussed above, pre-INPRES enrollment and INPRES school construction correlates at -0.13 (significant at the 1 % level), which means that such a "reduced form" estimation is quite "noisy". Still, it turns out that when we run this regression we find that pre-enrollment rates interacted with years since 1979 have the expected, statistically significant effect (results available upon request).

Table A2: Robustness: Controlling for interactions with socioeconomic covariates

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0173***	-0.0166***	-0.0160***	-0.0174***	-0.0181***	-0.0192***	-0.0156**							
	(0.00610)	(0.00617)	(0.00585)	(0.00611)	(0.00610)	(0.00645)	(0.00612)							
*(# Schools / # Children)$_i$ * Years since 1978$_t$*								-0.00305***	-0.00267***	-0.00349***	-0.00303***	-0.00304***	-0.00327***	-0.00271***
								(0.000705)	(0.000718)	(0.000766)	(0.000707)	(0.000705)	(0.000715)	(0.000810)
*School Attendance (5-14)$_i$ * Post-1978$_t$*	No	Yes	No	No	No	No	Yes	No	No	No	No	No	No	No
*Enrollment Rate$_i$ * Post-1978$_t$*	No	No	Yes	No	No	No	Yes	No	No	No	No	No	No	No
*Primary Industries$_i$ * Post-1978$_t$*	No	No	No	Yes	No	No	Yes	No	No	No	No	No	No	No
*Rural Population$_i$ * Post-1978$_t$*	No	No	No	No	Yes	No	Yes	No	No	No	No	No	No	No
*Employment Rate$_i$ * Post-1978$_t$*	No	No	No	No	No	Yes	Yes	No	No	No	No	No	No	No
*School Attendance (5-14)$_i$ * Years since 1978$_t$*	No	No	No	No	No	No	No	No	Yes	No	No	No	No	Yes
*Enrollment Rate$_i$ * Years since 1978$_t$*	No	No	No	No	No	No	No	No	No	Yes	No	No	No	Yes
*Primary Industries$_i$ * Years since 1978$_t$*	No	No	No	No	No	No	No	No	No	No	Yes	No	No	Yes
*Rural Population$_i$ * Years since 1978$_t$*	No	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes
*Employment Rate$_i$ * Years since 1978$_t$*	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.507	0.507	0.507	0.506	0.507	0.508
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The unit of observation is a district i. The variable *School Attendance (5-14)$_i$* represents the pre-INPRES share of school-aged children enrolled in school in district i. The variable *Enrollment Rate$_i$* represents the population-wise pre-INPRES enrollment rates observed in district i. The variable *Rural Population$_i$* represents the share of the population of district i living in rural areas. The variable *Primary Industries Employment$_i$* represents the share of the population working in primary industries (i.e., Agricultural and Mining Industries) observed in the district i. The variable *Employment Rate$_i$* represents the share of the population working observed in the district i. All socio-economic variables were computed using the 1971 Census (IPUMS (2018)). The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Robustness Analysis: Water and Sanitation Program

Below are displayed the results when controlling for the intensity of a water and sanitation program being implemented contemporaneously with INPRES. As discussed in section 6.3, the water and sanitation program does not impact the level of conflict, while the estimated impact of school construction remains virtually unchanged.

Table A3: Robustness: Controlling for the water and sanitation investment program

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)
(# Schools / # Children)$_i$ * *Post-1978*$_t$	-0.0138**	-0.0120*	-0.0187***			
	(0.00646)	(0.00683)	(0.00707)			
(# Schools / # Children)$_i$ * *Years Since 1978*$_t$				-0.00162***	-0.00213***	-0.00313***
				(0.000595)	(0.000801)	(0.000877)
Intensity Water and Sanitation Program$_i$ * *Post-1978*$_t$	0.00435	0.00783	0.00593			
	(0.0156)	(0.0137)	(0.0193)			
Intensity Water and Sanitation Program$_i$ * *Years since 1978*$_t$				0.000655	0.00152	0.000349
				(0.00155)	(0.00162)	(0.00215)
Observations	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.450	0.506	0.346	0.450	0.506
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	Yes	No	Yes	Yes
Province x Year FEs	No	No	Yes	No	No	Yes
Sample Mean	.08	.08	.08	.08	.08	.08

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Intensity Water and Sanitation Program*$_i$ represents the intensity of a water sanitation program implemented contemporaneously with INPRES in district i. The dummy *Post-1978*$_t$ takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978*$_t$ is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Robustness Analysis: Migration and Rural Population

As discussed in subsection 6.4, we extend here the set of controls of our baseline Table 2 to include measures of migration and rural population. In particular, we use the Indonesian population census of 1971, 1980 and 1990 (IPUMS (2018)) to build a (rough) time-varying variable of migration, defined as the share of population in a given district and year that has immigrated from another province (note that we do not have information on between-district migration). Drawing on the same raw data, we also build a second control variable, namely the share of the population in a given district and year living in a rural area. Given that we only have 3 census waves of data available, we need to heavily interpolate the data to build these two variables. In particular, for pre-1971 values we use the 1971 value, between 1971 and 1990 we use linear interpolation, drawing on the closest observable data points, while post-1990 we assign the 1990 value. The scarcity of data results in very noisy measures and warrants great caution in the interpretation of the results.[27]

Table A4 displays the results when including the migration and rural population measures in our baseline specifications. We find that the results remain very similar when controlling for these variables. These findings need to be interpreted with caution, given that the available data only permits us to construct quite rough proxies for these measures.

Table A4: Robustness: Controlling for time-varying migration and rural population

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(# Schools / # Children)$_i$ * *Post-1978*$_t$	-0.0173***	-0.0174***	-0.0173***	-0.0174***				
	(0.00610)	(0.00610)	(0.00611)	(0.00610)				
(# Schools / # Children)$_i$ * *Years Since 1978*$_t$					-0.00305***	-0.00306***	-0.00305***	-0.00306***
					(0.000705)	(0.000706)	(0.000708)	(0.000710)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506
Migration$_{it}$	No	Yes	No	Yes	No	Yes	No	Yes
Rural Population$_{it}$	No	No	Yes	Yes	No	No	Yes	Yes
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978*$_t$ takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978*$_t$ is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. *Migrations*$_{it}$ represents the share of population in a district i and year t having immigrated from another province. *Rural Population*$_{it}$ corresponds to the share of population in a district i and year t living in rural areas. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Time-varying measures of migration and rural population were computed using the Indonesian population census of 1971, 1980 and 1990 (IPUMS (2018)) (additional details are provided in the text). Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.5 Robustness Analysis: Climate and Oil Shocks

As discussed in section 6.5, in what follows we study whether a series of localized shocks could drive our findings. In particular, we shall investigate the role of climate shocks, such as precipitation or temperature shocks, as well as spikes in natural resource rents. The raw data on climate shocks is taken from Tollefsen et al. (2012). Further, Indonesia being a sizeable oil producer, we focus in terms of resource rents on oil revenues.[28] Drawing on Prio-Grid (Tollefsen et al. (2012)) data on oil presence in a given grid cell, we construct a time-invariant indicator of whether in a given district ever oil has been depleted, and interact this variable of oil presence with the current world oil price (from BP Statistical Review of World Energy Prices). The fact of using the "oil potential" rather than the (arguably more endogenous) actual "oil production" follows the identification strategy implemented in Berman et al. (2017). As an alternative

[27]There are 10 districts for which the match over time was problematic. In these cases, we used average values at the province level to construct the two measures. The results are essentially the same when we remove these districts.

[28]Indonesia contributes about 1 percent of world oil production (Company (2018)), which makes oil a sizeable sector of the Indonesian economy, and hence an important shock to control for in a robustness exercise. This being said, reassuringly for our identification strategy, Indonesia is typically a small enough producer to not be able to affect the world oil price.

approach we also control for a dummy capturing whether there is any employment or not in the oil sector in each district from the 1971 Census (IPUMS (2018)) interacted with current world oil prices. As shown in Tables A5 and A6, respectively, the estimated impact of school construction is very robust to controlling for these climatic and natural resource shocks.

Table A5: Robustness: Controlling for climate and oil rent shocks 1/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children) * Post-1978*	-0.0173***	-0.0174***	-0.0174***	-0.0187***				
	(0.00610)	(0.00609)	(0.00607)	(0.00616)				
*(# Schools / # Children) * Years Since 1978*					-0.00305***	-0.00305***	-0.00309***	-0.00294***
					(0.000705)	(0.000704)	(0.000703)	(0.000690)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.506	0.506	0.506	0.506	0.506	0.506	0.507	0.507
Precipitations$_{it}$	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Temperature$_{it}$	No	No	Yes	Yes	No	No	Yes	Yes
Oil [PRIO-Grid]$_i$ x *Oil Prices$_t$*	No	No	No	Yes	No	No	No	Yes
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. Time-varying measures of precipitations and temperature were obtained from the Prio-Grid Tollefsen et al. (2012) data. The variable *Oil [Prio-Grid]* takes a value of 1 if oil has been depleted in district i over the period. World oil prices were retrieved from the BP Statistical Review of World Energy Prices) (additional details are provided in the text). The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Robustness: Controlling for climate and oil rent shocks 2/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children) * Post-1978*	-0.0173***	-0.0174***	-0.0174***	-0.0178***				
	(0.00610)	(0.00609)	(0.00607)	(0.00614)				
*(# Schools / # Children) * Years Since 1978*					-0.00305***	-0.00305***	-0.00309***	-0.00304***
					(0.000705)	(0.000704)	(0.000703)	(0.000701)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.506	0.506	0.506	0.506	0.506	0.506	0.507	0.507
Precipitations$_{it}$	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Temperature$_{it}$	No	No	Yes	Yes	No	No	Yes	Yes
Oil [Census 1971]$_i$ x *Oil Prices$_t$*	No	No	No	Yes	No	No	No	Yes
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. Time-varying measures of precipitations and temperature were obtained from the Prio-Grid Tollefsen et al. (2012) data. The variable *Oil [Census 1971]* takes a value of 1 if there was any employment in the oil sector in the district i in the 1971 Census IPUMS (2018). World oil prices were retrieved from the BP Statistical Review of World Energy Prices) (additional details are provided in the text). The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.6 Robustness Analysis: Intensive Margin

As discussed in Section 6.7, we at present investigate whether education may not only affect the extensive margin of experiencing conflict or not, but also play a role for the intensive margin of conflict frequency. More specifically, in the Table A7 we define count measures of the number of days, weeks, resp. months in a year and district featuring newspaper articles in the Sydney Morning Herald referring to conflict according to our algorithm. No matter whether we slice the data in terms of days, weeks or months we continue to find that schooling reduces the scope for conflict frequency.

Table A7: Robustness: Intensive margin 1/2

Dep. Variable:	(1) *log(Days+1)$_{it}$*	(2) *log(Weeks+1)$_{it}$*	(3) *log(Months+1)$_{it}$*	(4) *log(Days+1)$_{it}$*	(5) *log(Weeks+1)$_{it}$*	(6) *log(Months+1)$_{it}$*
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0122**	-0.0135**	-0.0139***			
	(0.0493)	(0.0111)	(0.00378)			
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*				-0.00191	-0.00227*	-0.00250***
				(0.211)	(0.0515)	(0.00148)
Observations	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.688	0.681	0.654	0.688	0.681	0.654
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. OLS estimates are reported in all columns. The dependent variable is defined as the (log) number of days, weeks or months featuring newspaper articles in the *Sydney Morning Herald* referring to conflict events in district i in year t. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Robustness: Intensive margin 2/2

Dep. Variable:	(1) *log(Avg. Length +1)*	(2) *Inverse Hyperbolic*	(3) *log(Avg. Length +1)*	(4) *Inverse Hyperbolic*
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.128***	-0.139***		
	(0.0457)	(0.0496)		
*(# Schools / # Children))$_i$ * Years Since 1978$_t$*			-0.0269***	-0.0291***
			(0.00518)	(0.00562)
Observations	11,560	11,560	11,560	11,560
R-squared	0.505	0.506	0.506	0.507
District FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. OLS estimates are reported in all columns. In columns 1 and 3 [2 and 4], the dependent variable is defined as the (log) [inverse hyperbolic sine] average length of featuring newspaper articles in the *Sydney Morning Herald* referring to conflict events in district i in year t. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Another way to proxy for the intensive margin is to take into account the length of a given newspaper article. Bigger events (e.g. more intensive fighting) should on average result in longer articles than more minor incidents. In Table A8 we hence use the average length of the articles related to conflict events as a proxy of the intensity of the events in each district-year. We find a strong effect of education reducing the average length of conflict-related newspaper articles, consistent with a drop in fighting intensity.

In a nutshell, we find in the current Appendix section that while education has a large effect on the extensive margin of conflict, we also observe that schooling construction pushed down the intensive margin of conflict.

A.7 Robustness Analysis: Main Keywords

An important parameter for the construction of the conflict data is the set of keywords used, as discussed in Section 6.7. In Table A9 below are listed the baseline set of terms used to identify conflict-related sentences following the procedure described in Online Appendix B.1.

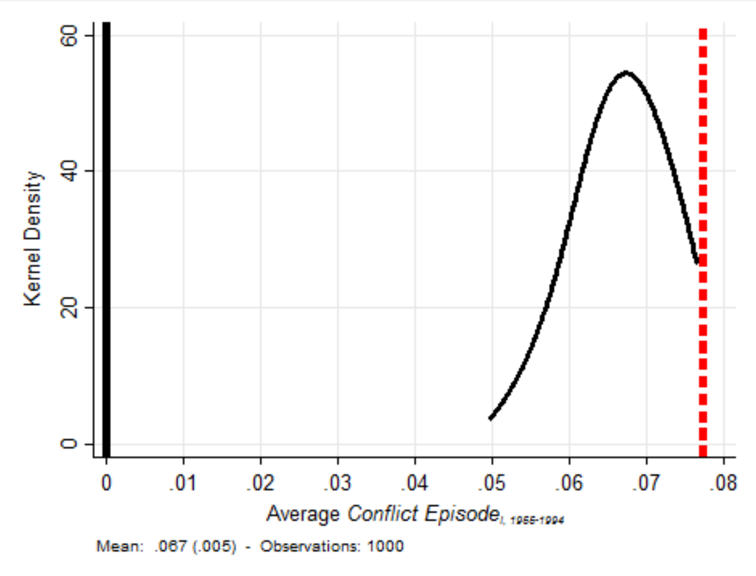
Table A9: Baseline Definition conflict keywords

Conflict, Battle, Assault, Kill, Riot, Attack, Turmoil, Unrest, Warfare, Soldier, Army, Insurgent, Terrorist, Disorder, Strike, Shoot, Massacre, Revolt.

Here we assess whether our results are robust when only a subsample of these keywords are used. To this end we carry out a Monte Carlo analysis with 1000 repetitions where for each draw only two-thirds of the baseline keywords are used. We first display in Figure 4 the distribution of the average conflict likelihood depending on the sample of keywords used. The dashed line represents the average number of conflict episodes obtained using the baseline set of keywords. While being by construction lower, the average number of conflict events obtained with a smaller number of keywords remains fairly stable and close to the one obtained using the whole set of conflict-related terms.

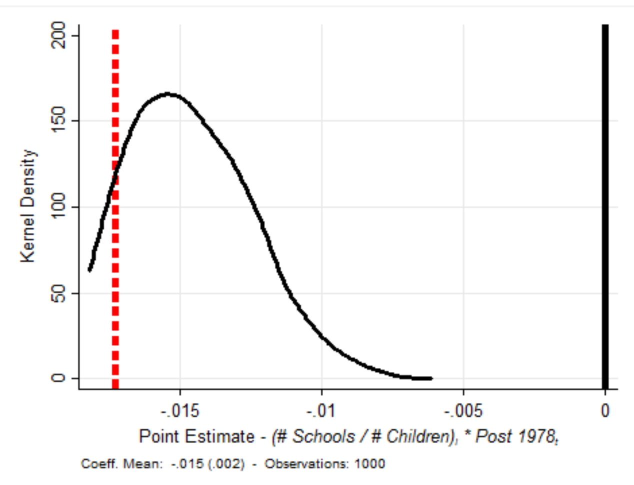
Further, we replicate our baseline results using each of the 1000 drawn conflict measures. Panels A and B in Figure 5 display the distribution of coefficients estimated using baseline regressions presented in columns 3 and 6 of Table 2, respectively. In both cases, point estimates of the coefficient of interest appear to be fairly stable and are consistent with the estimates reported in the main text.

Figure 4: Average conflict episodes estimated using 1,000 groups of conflict-related keywords

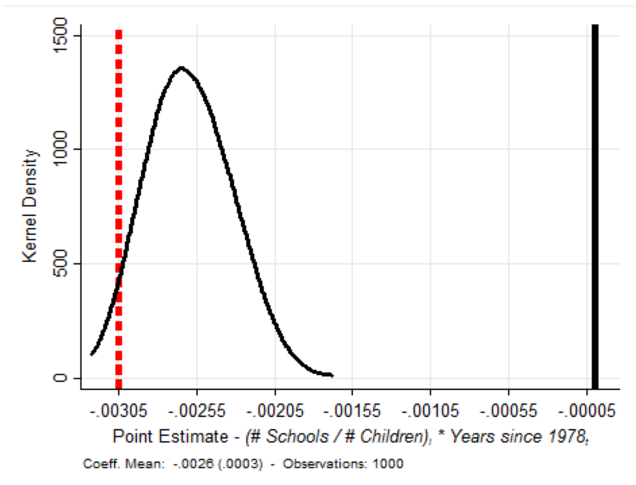


NOTE: The figure shows the distribution of the average conflict likelihood obtained using 1,000 different samples. Each sample was created by following the approach described in Section 4.1 and by randomly drawing only two-thirds of the baseline keywords. The dashed line represents the average number of conflict episodes obtained using in the full baseline set of keywords.

Figure 5: Distribution of coefficients estimated using 1,000 groups of conflict-related keywords



A) Column 3 in Table 2



B) Column 6 in Table 2

NOTE: The figure shows the distribution of coefficients estimated using 1,000 different samples. Each sample was created by following the approach described in Section 4.1 and by randomly drawing two thirds of the baseline keywords. Panel A and B depict the distribution of coefficients estimated using baseline regressions presented in columns 3 and 6 of Table 2, respectively. The dashed line represents the point estimate of the corresponding coefficient obtained using all baseline keywords.

A.8 Robustness Analysis: Broader Set of Keywords and Alternative Conflict Data

As discussed in Section 6.7, the current Appendix presents the findings when using a broader set of keywords, and another newspaper source, the *Canberra Times*. As far as a broader set of keywords are concerned, we now as a robustness check include the additional words listed in Table A10, which often refer to conflict events but may occasionally pick up "false positives".

Table A10: Broader Definition conflict keywords

Baseline Definitions + Engage, Defeat, Jar, Fight, Onslaught, Collide, Infringe, Onrush, Blast, Struggle, Upheaval, Hit, Combat, Tumult, Rebellion, Ravish, Forces, Slaughter, Assail, Guerrilla, Carnage, Snipe, Rebel, Uprising, Blash, Insurrection, Butchery, Aggression, Terrorism, Clash, Smash.

Concerning the alternative newspaper source, as explained in Online Appendix B.1, both in terms of readership and uninterrupted coverage the *Sydney Morning Herald* is preferable to the *Canberra Times*. Still, considering a second media source is useful. Table A11 below displays the results when replicating the columns 1 and 3 of our baseline Table 2 using broader keywords or *Canberra Times* articles. In particular, columns 1 and 2 of Table A11 reproduce the columns 1 and 3 of our baseline Table 2, while in columns 3-4 of Table A11 the aforementioned set of keywords is used of the data construction, in columns 5-6 the *Canberra Times* is used as sole source of information instead of the *Sydney Morning Herald*, while in columns 7-8 a district year is coded as having conflict if this has been featured in an article of either the *Canberra Times* or the *Sydney Morning Herald*. Table A12 performs the analogous robustness exercises but for the columns 4 and 6 of Table 2 (instead of columns 1 and 3). The results of Tables A11 and A12 point out that our findings are similar when broadening the keywords considered or newspaper source adopted.

Table A11: Robustness: Alternative sources 1/2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable: Conflict Episode$_{it}$	*SMH*	*SMH*	*SMH-Broad*	*SMH-Broad*	*CT*	*CT*	*SMH+CT*	*SMH+CT*
(# Schools / # Children)$_i$ * *Post-1978*$_t$	-0.0127***	-0.0173***	-0.00912*	-0.00988	-0.00710*	0.000292	-0.0161***	-0.0197***
	(0.00448)	(0.00610)	(0.00496)	(0.00847)	(0.00388)	(0.00564)	(0.00474)	(0.00677)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.506	0.375	0.544	0.325	0.460	0.392	0.545
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes	No	Yes
Sample Mean	.08	.08	.13	.13	.06	.06	.11	.11

NOTE: The unit of observation is a district i and year t. The dataset covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t. The variable *# Schools/# Children*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978*$_t$ takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). columns 1 and 2 report results obtained using conflict data constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. In columns 3 and 4, conflict data were obtained using the *Sydney Morning Herald* and a more coarse set of conflict-related keywords (see discussion in Appendix A.7). Columns 5 and 6 report results obtained using conflict data constructed following the approach described in Section 4.1 using the *Canberra Times*. In columns 7 and 8 the dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t when we combine the two sources. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Robustness: Alternative sources 2/2

Dep. Variable: Conflict Episode$_{it}$	(1) *SMH*	(2) *SMH*	(3) *SMH-Broad*	(4) *SMH-Broad*	(5) *CT*	(6) *CT*	(7) *SMH+CT*	(8) *SMH+CT*
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.00146***	-0.00305***	-0.00133***	-0.00256***	-0.000882**	-0.00118**	-0.00176***	-0.00321***
	(0.000421)	(0.000705)	(0.000449)	(0.000856)	(0.000348)	(0.000512)	(0.000432)	(0.000786)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.346	0.506	0.375	0.545	0.325	0.460	0.392	0.546
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes	No	Yes
Sample Mean	.08	.08	.13	.13	.06	.06	.06	.06

NOTE: The unit of observation is a district i and year t. The dataset covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t. The variable *# Schools/# Children$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. columns 1 and 2 report results obtained using conflict data constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. In columns 3 and 4, conflict data were obtained using the *Sydney Morning Herald* and a more coarse set of conflict-related keywords (see discussion in Appendix A.7). Columns 5 and 6 report results obtained using conflict data constructed following the approach described in Section 4.1 using the *Canberra Times*. In columns 7 and 8 the dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t when we combine the two sources. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.9 Robustness Analysis: Alternative Conflict Data

The coverage period of our conflict measure constructed using the Sydney Morning Herald (abbreviated, SMH) overlaps from 1979-1994 with the conflict measure from GDELT (GDELT (2018)) that covers 1979-2014, but there is no temporal overlap with the conflict variable from ICEWS (ICEWS (2018)), as ICEWS starts in 1995 (and finishes in 2014) and SMH data ends in 1994.[29] Similarly, there is also lack of overlap with the NVMS (NVMS (2019)) data. The coverage of NVMS starts in 1998 for nine conflict-prone provinces and increases to 15 provinces plus greater Jakarta beginning in 2005, but the data is not representative of Indonesia and the coverage is judged less reliable for the earliest years (see Bazzi and Gudgeon (2018)).[30] Hence, we use NVMS data from 2005 to 2014.

Still, the overlap allows us to compare SMH with GDELT, revealing that for 86 percent of observations these two variables have the same values (i.e. both 0 or both 1). The parallel evolution over time of our measure when compared with GDELT is displayed graphically in Figure 6. The correspondence for SMH and GDELT seems relatively high in the light of the fact that for the period where ICEWS and GDELT overlap, they take on the same values for only 62 percent of observations. Put differently, our measure is much more similar to GDELT than is the case for ICEWS which differs substantially from GDELT. When comparing GDELT with NVMS, we observe an increasing level of correspondence over time, as NVMS becomes more representative (i.e. covers more of Indonesia) and becomes arguably more precise. While for the period (2005-2014) these measures only correspond in 66 percent of cases, they have 78 percent of correspondence for the end of this period (i.e. for 2011-2014).

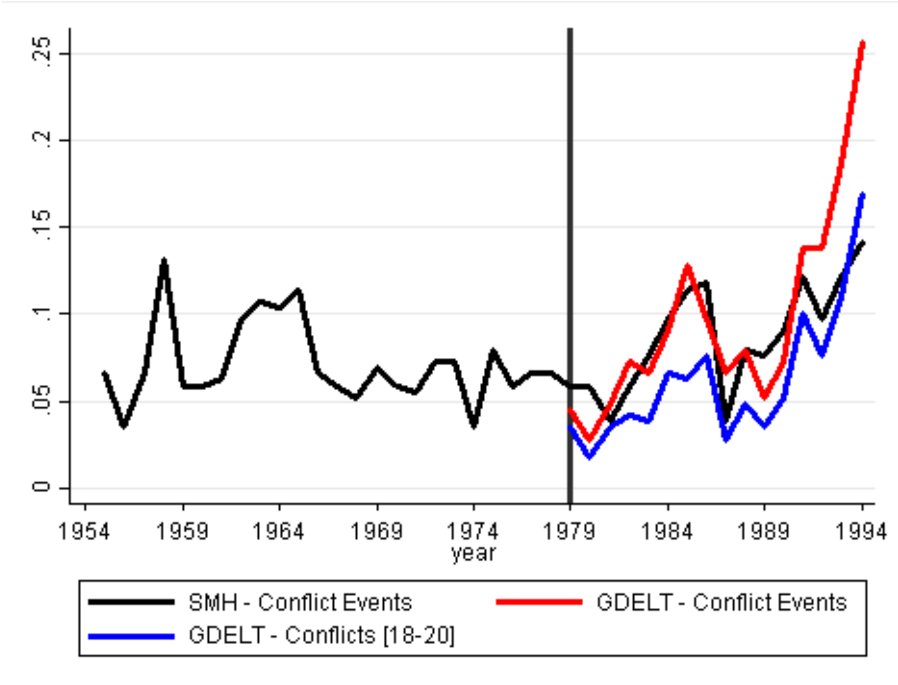
As a next step, below we use these three existing datasets on conflict in Indonesia to assess the robustness of our results. All of these datasets, GDELT (GDELT (2018)), ICEWS (ICEWS (2018)) and NVMS (NVMS (2019)), have the downside of only covering a time period after the INPRES school construction program, which rules out the difference-in-difference analysis that we carry out in our baseline regressions. In particular, our data from GDELT covers 1979-2014, the ICEWS sample stretches over 1995-2014, and we use NVMS for the period 2005-2014, allowing us to obtain a balance panel of conflict episodes in 194 districts (out of 289). However, the analysis of the increasing impact of school construction (see e.g. the columns 4-6 of the baseline Table 2) can be replicated using the GDELT, ICEWS and NVMS data, which is what we do below in the Tables A13, A14, and A15 respectively. Reassuringly, in all these tables we find comparable results as in our main analysis. The quantitatively smaller size of the

[29] For the main GDELT and ICEWS measures we focus on their categories 15 to 20 of events to code them as "conflict", and narrow as robustness check the definition of conflict down to containing only their categories 18 to 20.

[30] Note that the focus of NVMS on conflict-prone provinces and on including also lower-scale events means that the unconditional probability of observing a conflict event in our sample is close to 0.85, which implies that the data only offers limited identifying variation.

coefficients for the NVMS estimates is consistent with the expectation that the increasing effect of school construction flattens out after some years.

Figure 6: Evolution of conflict episodes across alternative sources



SOURCE: Authors' computations from GDELT (2018) and own conflict data. SMH conflict data is obtained using the procedure described in Section 4.1.

Table A13: Robustness: Conflict data from GDELT (2018)

	All Conflict Events$_{it}$				Conflicts [18-20]$_{it}$			
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.000311	-0.00129**	-0.000965	-0.00201***	-0.000454	-0.00137**	-0.000919	-0.00194***
	(0.000883)	(0.000655)	(0.000971)	(0.000628)	(0.000605)	(0.000585)	(0.000677)	(0.000575)
Observations	4,624	10,404	4,624	10,404	4,624	10,404	4,624	10,404
R-squared	0.353	0.574	0.419	0.624	0.325	0.527	0.388	0.583
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	No	No	Yes	Yes	No	No	Yes	Yes
Time-Window	1979-1994	1979-2014	1979-1994	1979-2014	1979-1994	1979-2014	1979-1994	1979-2014
Sample Mean	.1	.41	.1	.41	.06	.27	.06	.27

NOTE: The unit of observation is a district i and year t. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t. The full dataset covers 289 districts across 26 provinces over the period 1979-2014. LPM estimates are reported in all columns. Conflict data from GDELT (2018). In the first (last) four columns we code the categories 15 to 20 (18 to 20) of events as *conflict*. The variable *# Schools/# Children$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Robustness: Conflict data from ICEWS (2018)

	All Conflict Episodes$_{it}$		Conflicts [18-20]$_{it}$	
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.00129*	-0.00161*	-0.00103*	-0.000789
	(0.000721)	(0.000960)	(0.000604)	(0.000731)
Observations	5,780	5,780	5,780	5,780
R-squared	0.436	0.501	0.360	0.434
District FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Province x Year FEs	No	Yes	No	Yes
Time-Window	1995-2014	1995-2014	1995-2014	1995-2014
Sample Mean	.4	.4	.25	.25

NOTE: The unit of observation is a district i and year t. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t. The dataset covers 289 districts across 26 provinces over the period 1995-2014. LPM estimates are reported in all columns. Conflict data from ICEWS (2018). In the first (last) two columns we code the categories 15 to 20 (18 to 20) of events as *conflict*. The variable *# Schools/# Children$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Robustness: Conflict data from NVMS (2019)

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.00323**	-0.00149
	(0.00136)	(0.000998)
Observations	1,800	1,800
R-squared	0.695	0.788
District FEs	Yes	Yes
Year FEs	Yes	Yes
Province x Year FEs	No	Yes
Time-Window	2005-2014	2005-2014
Sample Mean	.85	.85

NOTE: The unit of observation is a district i and year t. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i in year t. The dataset covers 194 districts across 17 provinces over the period 1995-2014. LPM estimates are reported in all columns. Conflict data from **?**. The variable *# Schools/# Children$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ONLINE APPENDIX
Education and Conflict
Evidence from a Policy Experiment in Indonesia

Dominic Rohner[†] Alessandro Saia[‡]

March 15, 2019

[*]First version: December 2018. Acknowledgements: We are grateful for helpful comments from Siwan Anderson, Chris Blattman, Michael Böhm, Patrick Francois, Luis Martinez, Munir Squires, Felipe Valencia Caicedo, and seminar audiences at Chinese University of Hong Kong, Hong Kong University, Simon Fraser University, Tilburg University, University of British Columbia, University of Chicago. Dominic Rohner and Alessandro Saia gratefully acknowledge financial support from the ERC Starting Grant POLICIES FOR PEACE-677595. We thank Esther Duflo for sharing her data with us.

[†]Department of Economics, University of Lausanne and CEPR. E-mail: dominic.rohner@unil.ch.

[‡]Department of Economics, University of Lausanne. E-mail: alessandro.saia@unil.ch.

In the Online Appendices below we provide additional investigation and further results for the various sections of the paper. We shall always start the title of a given Online Appendix section with the same wording as the corresponding section in the main text. For example, the first Online Appendix section B.1 labeled "Data: Construction of the conflict measure" provides additional details with respect to the section called "Data" (Section 4) in the main text.

Below is listed the Table of Content of the Online Appendices.

B.1 Data: Construction of the conflict measure

As briefly summarized in Section 4 of the main text, our approach to construct a novel geo-referenced dataset of conflict-related events in Indonesia consists of five steps.

The *first step* was to identify a valid source of data, i.e. a newspaper with a historical digital archive allowing to cover our sample period 1955-1994. Complete time coverage of such a long period starting more than 60 years ago is very rare, but thankfully we have found a high-quality outlet, namely the Sydney Morning Herald (thereafter, SMH) which is a daily newspaper published by Fairfax Media in Sydney, Australia.[1] Founded in 1831, the SMH is the oldest continuously published newspaper in Australia and currently has a readership of roughly half a million people (Morgan (2018)). According to Media Bias/Fact Check (https://mediabiasfactcheck.com), the SMH has a slight to moderate liberal bias with high quality of factual reporting. Using a major newspaper that is based in Australia has the advantage of being geographically quite close to Indonesia without suffering from obvious political biases in reporting. The SMH digital archive provides full-digital text coverage to every edition of the newspaper published between January 1st, 1955 and February 2nd, 1995. As a consequence, we have been able to construct a database of violent events in Indonesia between 1955 and 1994.

After having identified the newspaper, the *second step* was to analyze the underling unstructured text data (i.e. newspaper articles) to construct the desired information. To this end, we have performed a first selection of the articles where we retrieved all SMH articles related to Indonesia. In particular, we searched over 820,000 articles available in the SMH archive and downloaded all those containing at least once the word "Indonesia" (the resulting set of articles was of around 34,000).

In a *third step*, we used natural language processing algorithms to analyse the content of all 34,000 articles. In doing so, we screened all sentences contained in all 34,000 articles and extracted all sentences where at least one conflict related term was present.[2] Concretely, we divided all articles in sentences and searched sentence by sentence for a conflict-related term. If a term was found, we stored the sentence for use in the following step.

Then (*fourth step*), we used a Named Entity Recognition algorithm to identify all real world entities contained in all tagged sentences (i.e. all real-world objects that can be denoted with a proper name and have a physical existence). A named entity is a real-world object, such

[1] There exist also some other newspapers with digital archives, such as e.g. the New York Times, but they typically have major restrictions on the number of articles downloadable per month, making the data collection over such a large sample period with dozens of thousands of articles impracticable.

[2] The conflict-related terms used in the main analysis were: "conflict" "battle" "assault" "kill" "riot" "attack" "turmoil" "unrest" "warfare" "solider" "army" "insurgent" "terrorist" "disorder" "revolt" "massacre" "strike" plus all their variations (i.e. the terms with the suffix "ing" / "s" / "es" or "ed"). We also use as robustness checks i) a sub-set of our main keywords, and ii) a larger set of keywords (see Appendices A.7 and A.8, respectively).

as people, locations or organizations. As a consequence, not all the entities identified were locations. Thus, we performed a *fifth (and final) step* where we matched all entities with locations contained in a digital gazetteer of geographical entities in Indonesia.

After this final step, we were able to identify both the geographical coordinates of matched locations (and in which district (kabupate) they were located) and the time of the event (i.e. the date of the article).

As discussed also in the main text, we have performed a wide set of robustness exercises to assess the validity of our conflict measure. The main sensitivity tests were developed along four different dimensions.

First, we perform all the above mentioned steps using a second newspaper: the Canberra Times (thereafter, CT) which is another Australian newspaper with a digital archive available over the period of interest. We prefer to rely on the SMH as main information source and use the CT as source for robustness checks, as first the CT is much smaller (with its readership being about a tenth of the one of SMH, according to Roy Morgan Research), and second its archive does not contain all issues (e.g. in 1955 there are 347 issues available, in 1965 the number is 331, in 1975 322 and in 1985 366). This being said, reassuringly, we find similar results when replicating our analysis using CT (see Appendix A.8).

Second, we have also performed additional exercises finding that our results hold when alternative matching scores are adopted in the matching of entities and locations in the gazetteer (i.e. in the fifth step, rather than using the perfect match we adopted a fuzzy match) (see Online Appendix B.10).

Third, we replicate below our analysis using an alternative python algorithm to identify locations. In particular, in the main analysis we used the Stanford Named Entities tagger present in the NLTK module, while in this robustness test we relied on the *geotext* module, which appeared to be "faster" then the NLTK module but less accurate (especially when locating entities such as areas or regions). Reassuringly, results obtained using this alternative algorithm support the main findings of the paper (see Online Appendix B.11).

Finally, we have used three alternative conflict databases to replicate our analysis, as discussed below. These existing databases available for Indonesia are GDELT (2018), ICEWS (2018) and NVMS (2019). They entail the downside of barring us from performing a difference-in-difference analysis, as they do not cover the period prior to the INPRES program (GDELT starts in 1979, ICEWS begins only in 1995, and NVMS starts (partial) coverage in 1998). Therefore, when using these alternative data sources we are only able to perform an empirical exercise where the identification strategy relies on the variation over time of the effect of the program. Also in this case the results obtained with these alternative sets of conflict data are consistent with the findings presented in the main analysis (see Appendix A.9).

B.2 Data: Additional Descriptive Statistics

Table B1: Additional descriptive statistics

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
INDONESIAN POPULATION CENSUS OF 1971$_i$:					
Enrollment Population$_i$	0.178	0.096	0.031	0.993	289
School Attendance (5-14)$_i$	0.483	0.154	0.016	0.841	289
Rural Population$_i$	0.736	0.382	0	1	289
Primary Industries Employment$_i$	0.605	0.308	0	0.993	289
Mining Employment$_i$	0.004	0.027	0	0.335	289
Agricultural Employment$_i$	0.6	0.31	0	0.99	289
Religious Polarization$_i$	0.23	0.309	0	0.998	289
Return to Education [Bricks]$_i$	-0.511	0.498	-4.180	0	238
Return to Education [Entrep.]$_i$	-1.062	1.67	-17.238	0	227
Average RoE Bricks and Entrep.$_i$	-0.772	1.116	-13.845	0	262
Princ. Comp. RoE Bricks and Entrep.$_i$	0.003	1.027	-7.047	1.232	203

SOURCE: Authors' computations from the Indonesian population census of 1971 (IPUMS (2018)).

Table B2: Cross-correlations

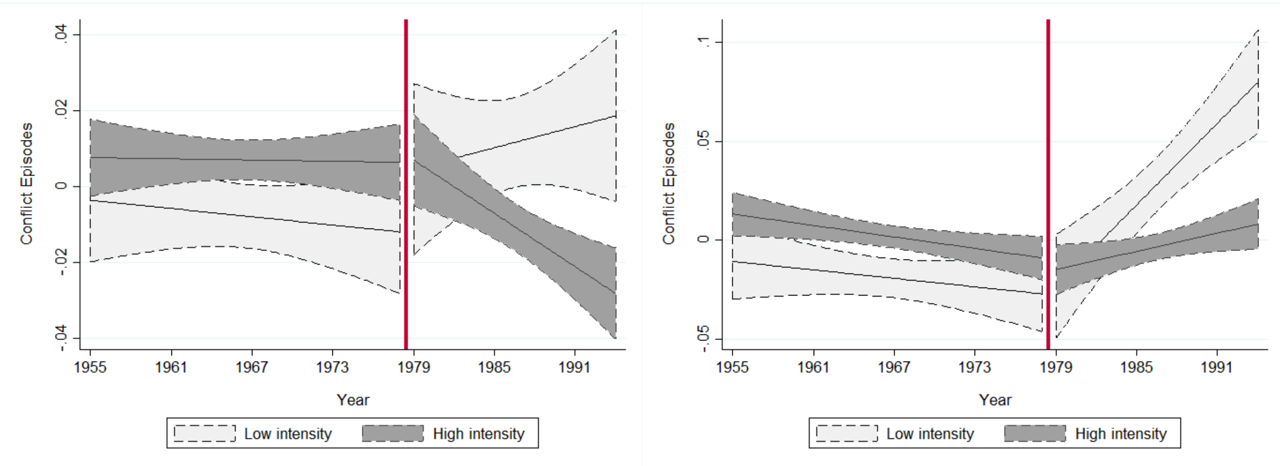
	# INPRES$_i$	Children 5-14$_i$	Enrol. Pop.$_i$	School Att.$_i$	Rural Pop.$_i$	Prim. Ind.$_i$	Mining$_i$	Agric.$_i$	Years Confl.$_i$	Dummy Confl.$_i$
# INPRES Schools$_i$	1.000									
Children 5-14$_i$	0.865	1.000								
Enrollment Population$_i$	-0.254	-0.213	1.000							
School Attendance$_i$	-0.315	-0.186	0.353	1.000						
Rural Population$_i$	0.376	0.281	-0.269	-0.492	1.000					
Primary Industry Employment.$_i$	0.179	0.042	-0.196	-0.447	0.805	1.000				
Mining Employment$_i$	-0.078	-0.071	-0.007	0.095	-0.094	-0.008	1.000			
Agricultural Employment$_i$	0.185	0.048	-0.195	-0.453	0.810	0.996	-0.094	1.000		
Years with Conflict [Pre-1979]$_i$	0.113	0.163	0.116	0.041	0.047	0.030	-0.058	0.034	1.000	
Dummy Conflict [Pre-1979]$_i$	0.056	0.114	0.072	0.077	0.060	0.028	-0.086	0.036	0.539	1.000

SOURCE: Authors' computations from Duflo (2001) and the Indonesian population census of 1971 (IPUMS (2018)). Conflict data were computed using the procedure described in Section 4.1.

B.3 Empirical Strategy and Main Results: Additional Common Trend Figures

As discussed in Section 5.2 of the main text, we provide below additional visual representations of the common pre-trend before the INPRES school construction treatment. It turns out that no matter how the sample is sliced in terms of school construction intensity, the assumption of a common pre-trend appears reasonable. Note that to fully guarantee a common pre-trend we have also implemented a synthetic control group approach in section 6.1, which by construction creates a parallel pre-trend for the synthetic control group. All our results are robust to this alternative exercise.

Figure B1: Additional common trend figures



A) 33%*ile* vs 66%*ile* Demeaned district-year B) 25%*ile* vs 75%*ile* Demeaned district

Source: Authors' computations from Duflo (2001) and own conflict data. Conflict data is obtained using the procedure described in Section 4.1. The figure shows the linear prediction plot with confidence intervals of normalized conflict events over pre-1978 and post-1978 periods (i.e., the year when INPRES program is completed). Left panel: Normalized conflict events in a district-year are computed by removing the sample mean of conflict episodes observed in the whole sample in the corresponding year, as well as the district mean over time. Low intensity and high intensity indicate areas with more versus less INPRES school construction, respectively. Low (high) areas are defined as all districts where the number of schools is below the 33th percentile (above the 66th percentile). Right panel: Normalized conflict events in a district-year are computed by removing the district mean of conflict episodes. Low intensity and high intensity indicate areas with more versus less INPRES school construction, respectively. Low (high) areas are defined as all districts where the number of schools is below the 25th percentile (above the 75th percentile).

B.4 Robustness Analysis: Alternative levels of clustering

As mentioned in the Section 6.6 of the main text, in the two Tables B3 and B4 below we show that the conclusions of the statistical inference continue to hold when we allow for standard errors to be clustered at alternative levels. In particular, Table B3 allows for standard errors to be clustered at the level of the 26 Indonesian provinces (although this number of clusters is arguably below the conventional minimum levels used in the literature), while Table B4 allows for standard errors to be two-way clustered at the district and year levels. In both cases of Tables B3 and B4 the coefficients of interest remain statistically significant.

Table B3: Clustering of standard errors at the Province level

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
# Schools / # Children * Post 1978	-0.0127***	-0.0101**	-0.0173***						
	(0.00297)	(0.00480)	(0.00514)						
# Schools / # Children * Years Since 1978				-0.00146***	-0.00175***	-0.00305***			
				(0.000312)	(0.000429)	(0.000818)			
# Schools / # Children [1979-1984]							-0.00658**	-0.00952**	-0.0168***
							(0.00248)	(0.00433)	(0.00479)
# Schools / # Children [1985-1989]							-0.0111***	-0.0151**	-0.0253***
							(0.00395)	(0.00680)	(0.00700)
# Schools / # Children [1990-1994]							-0.0218***	-0.0268***	-0.0489***
							(0.00451)	(0.00663)	(0.0101)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.450	0.506	0.346	0.450	0.506	0.346	0.450	0.507
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Province x Year FEs	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.08	.08	.08	.08	.08	.08	.08	.08	.08

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978*$_t$ takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978*$_t$ is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Years 1979-1984*$_t$ is a dummy taking a value of 1 for the years 1979-1984 (it is analogous for the two variables referring to the period 1985-1989 and 1990-1994, respectively). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the province level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Two-way clustering of standard errors at the district and year level

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
# Schools / # Children * Post 1978	-0.0127**	-0.0101	-0.0173**						
	(0.00508)	(0.00741)	(0.00834)						
# Schools / # Children * Years Since 1978				-0.00146***	-0.00175**	-0.00305***			
				(0.000438)	(0.000715)	(0.000843)			
# Schools / # Children [1979-1984]							-0.00658	-0.00952	-0.0168**
							(0.00452)	(0.00665)	(0.00707)
# Schools / # Children [1985-1989]							-0.0111*	-0.0151	-0.0253**
							(0.00645)	(0.0102)	(0.0111)
# Schools / # Children [1990-1994]							-0.0218***	-0.0268**	-0.0489***
							(0.00607)	(0.0117)	(0.0138)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.450	0.506	0.346	0.450	0.506	0.346	0.450	0.507
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Province x Year FEs	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.08	.08	.08	.08	.08	.08	.08	.08	.08

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978*$_t$ takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978*$_t$ is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Years 1979-1984*$_t$ is a dummy taking a value of 1 for the years 1979-1984 (it is analogous for the two variables referring to the period 1985-1989 and 1990-1994, respectively). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard errors two-way clustered at the district and year levels are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.5 Robustness Analysis: Logit

As discussed in the main text in Section 6.6, in the current Online Appendix section we will replicate our main baseline specifications using conditional logit regressions instead of the linear probability model that we apply throughout the paper.[3] As for the baseline analysis, we find a statistically significant conflict-reducing effect of the number of INPRES schools constructed.

Table B5: Impact of INPRES school construction on conflict: Fixed effects logit estimator

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)
(# Schools / # Children)$_i$ Post-1978$_t$*	-0.435**	-0.970***		
	(0.209)	(0.307)		
*(# Schools / # Children)$_i$ * Years since 1978$_t$*			-0.0323*	-0.0783**
			(0.0169)	(0.0374)
Pseudo R-squared	0.281	0.386	0.280	0.386
Observations	5,920	5,920	5,920	5,920
District FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes
Sample Mean	.14	.14	.14	.14

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. Fixed effects logit estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[3] Note that when including province-year fixed effects, the estimator does not converge. Hence, we restrict ourselves to the inclusion of district fixed effects, year fixed effects and district-specific time trends.

B.6 Robustness Analysis: Province Level Results

Below we collapse the sample at a larger level of aggregation, building namely a panel at the province-year level. As discussed in Section 6.6 of the main text, the find the same robust conflict-reducing impact of school construction at this larger level of aggregation.

Table B6: Robustness: Province level

	Dummy Conflict$_{pt}$		(log) Districts with Conflict$_{pt}$	
Dep. Variable:	(1)	(2)	(3)	(4)
*(# Schools / # Children)$_p$ * Post-1978$_t$*	-0.0434**		-0.0695***	
	(0.0220)		(0.0219)	
*(# Schools / # Children)$_p$ * Years Since 1978$_t$*		-0.00432**		-0.00777***
		(0.00212)		(0.00215)
Observations	1,040	1,040	1,040	1,040
R-squared	0.620	0.620	0.679	0.680
Province FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Sample Mean	.45	.45	.4511	.4511

NOTE: The unit of observation is a province p and year t. The sample covers 26 provinces over the period 1955-1994. LPM (OLS) estimates are reported in the first (last) two columns. In columns 1 and 2 the dependent variable is a dummy that takes a value of 1 if a violent event was observed in province p and year t. In columns 3 and 4 the dependent variable is the share of districts of province p and year t with violent events. The variable *(# Schools/# Children)$_p$* represents the average number of primary schools constructed under the INPRES program per 1,000 school-aged children in all districts of province p. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.7 Robustness Analysis: Alternative Sample Subperiods

As discussed in Section 6.6, in the current Online Appendix we replicate the results of columns 7-9 of our baseline Table 2 when splitting the sample period in a different number of subperiods. The results are displayed below in Table B7. We continue to find that the pacifying effect of education is increasing over time.

Table B7: Robustness: Alternative sample subperiods

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
*(# Schools / # Children)$_i$ * Years 1979-1986$_t$*	-0.0100**	-0.00973*	-0.0164***									
	(0.00444)	(0.00560)	(0.00600)									
*(# Schools / # Children)$_i$ * Years 1979-1994$_t$*	-0.0155***	-0.0150*	-0.0294***									
	(0.00511)	(0.00856)	(0.00879)									
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$*				-0.00658	-0.00952*	-0.0168***						
				(0.00427)	(0.00536)	(0.00591)						
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$*				-0.0111**	-0.0151*	-0.0253***						
				(0.00540)	(0.00813)	(0.00837)						
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$*				-0.0218***	-0.0268***	-0.0489***						
				(0.00594)	(0.0101)	(0.0110)						
*(# Schools / # Children)$_i$ * Years 1979-1982$_t$*							-0.00249	-0.00520	-0.0149**			
							(0.00439)	(0.00542)	(0.00623)			
*(# Schools / # Children)$_i$ * Years 1983-1986$_t$*							-0.0176***	-0.0211***	-0.0268***			
							(0.00598)	(0.00750)	(0.00762)			
*(# Schools / # Children)$_i$ * Years 1987-1990$_t$*							-0.00739	-0.0117	-0.0208**			
							(0.00510)	(0.00846)	(0.00855)			
*(# Schools / # Children)$_i$ * Years 1991-1994$_t$*							-0.0235***	-0.0286***	-0.0513***			
							(0.00635)	(0.0104)	(0.0115)			
*(# Schools / # Children)$_i$ * Years 1979-1981$_t$*										-0.00318	-0.00627	-0.0144**
										(0.00476)	(0.00550)	(0.00587)
*(# Schools / # Children)$_i$ * Years 1982-1984$_t$*										-0.00997*	-0.0137**	-0.0200**
										(0.00512)	(0.00652)	(0.00773)
*(# Schools / # Children)$_i$ * Years 1985-1987$_t$*										-0.0123**	-0.0168*	-0.0283***
										(0.00623)	(0.00852)	(0.00907)
*(# Schools / # Children)$_i$ * Years 1988-1990$_t$*										-0.0111**	-0.0163*	-0.0263***
										(0.00547)	(0.00901)	(0.00922)
*(# Schools / # Children)$_i$ * Years 1991-1994$_t$*										-0.0235***	-0.0295***	-0.0534***
										(0.00635)	(0.0106)	(0.0116)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.345	0.450	0.506	0.346	0.450	0.507	0.346	0.450	0.507	0.346	0.450	0.507
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Province x Year FEs	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years 1979-1984$_t$* is a dummy taking a value of 1 for the years 1979-1984 (it is analogous for the other variables referring to the other subperiods). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.8 Robustness Analysis: Exploiting annual INPRES school construction numbers

As mentioned in Section 6.6, below we make use of the information on the annual INPRES construction numbers. In particular, rather than applying 1979 as first year of the treatment (i.e. five years after 1974) for all districts, we rely in the current Online Appendix section on two alternative district-specific starting dates for our treatment. First, we focus on "*Year Max + 5*", where "Year Max" corresponds to the year with the highest number of INPRES schools constructed in the district (i.e. the mode). So if, say, e.g. in all years 0 or 1 schools get constructed, but in 1975 two new schools were built, then "*Year Max + 5*" would be 1980 (i.e. five years after the mode year). The second variant we consider is "*Year Half + 5*", where "*Year Half*" corresponds to when at least half of INPRES schools where constructed in the district. Again, if e.g. in total 4 schools were built in a district over the entire INPRES period and the second school was finished in 1976 when "*Year Half + 5*" would take a value of 1981. Notice that for these two exercises the sample is slightly smaller than in our baseline Table 2 (i.e 11,280 vs 11,560), as for some districts, we only know the total number of schools constructed but not the annual construction numbers.

The results of these robustness checks are displayed in Table B8. Reassuringly, the findings of these robustness checks are very similar to the main results of our baseline Table 2.

Table B8: Robustness: Exploiting Annual School Construction Numbers

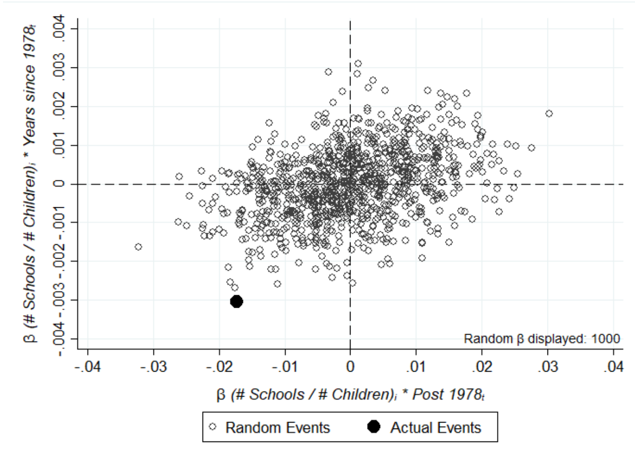
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Post-Year Max + 5$_t$*	-0.0148***	-0.0128***						
	(0.00462)	(0.00469)						
*(# Schools / # Children)$_i$ * Years since Year Max + 5$_t$*			-0.000972***	-0.00133***				
			(0.000288)	(0.000358)				
*(# Schools / # Children)$_i$ * Post-Year Half + 5$_t$*					-0.0141***	-0.0147***		
					(0.00469)	(0.00433)		
*(# Schools / # Children)$_i$ * Years since Year Half + 5$_t$*							-0.000920***	-0.00141***
							(0.000288)	(0.000325)
Observations	11,280	11,280	11,280	11,280	11,280	11,280	11,280	11,280
R-squared	0.344	0.505	0.344	0.505	0.343	0.505	0.344	0.505
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes	No	Yes
Sample Mean	.08	.08	.08	.08	.08	.08	.08	.08

NOTE: The unit of observation is a district i and year t. The sample covers 282 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The further variables included are described in the text of Online Appendix B.8. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

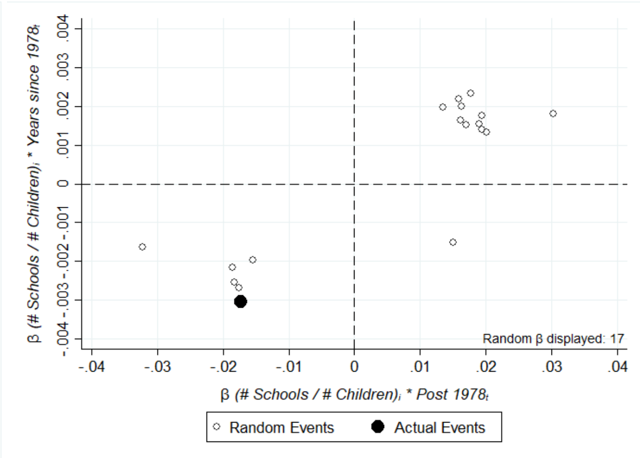
B.9 Robustness Analysis: Placebo

As discussed in Section 6.7 of the main text, to investigate concerns about our main findings having been obtained "by chance", we carry out a placebo exercise where we randomly assign treatment in 1000 placebo datasets with the same average conflict likelihood as the "true" data (i.e. our main conflict dataset built based on SMH articles). Figure B2 below depicts the clouds of estimated coefficients of our baseline specifications (Columns 3 and 6 of baseline Table 2) with this "fake" data. Panel A displays all coefficients obtained from all 1000 placebo samples. Each dot corresponds to one combination of coefficients in a cartesian plane where the horizontal axis represents the beta coefficient of the specification of Column 3, while the vertical axis depicts the beta coefficient of the specification of Column 6. The large black dot represents our true coefficients. Panel B shows the estimates when the coefficients obtained with the two specifications (and the same placebo dataset) are both statistically significant at the 10 % level: there are only 17 placebo datasets (out of 1000) for which we obtain statistically significant results using the two models. If in Panel C we use the 5 % significance threshold, the number of placebo datasets that satisfy this criterion is of 4 (out of 1,000). Finally, when in Panel D we use the 1 % significance level (which is the level of statistical significance obtained using the true data) there are no placebo datasets that satisfy this criterion. These results highlight how extremely unlikely it would have been to obtain our results "by chance".

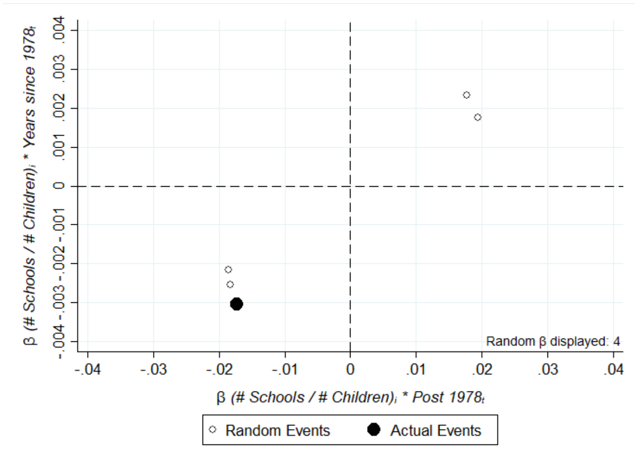
Figure B2: Results of Placebo Exercise



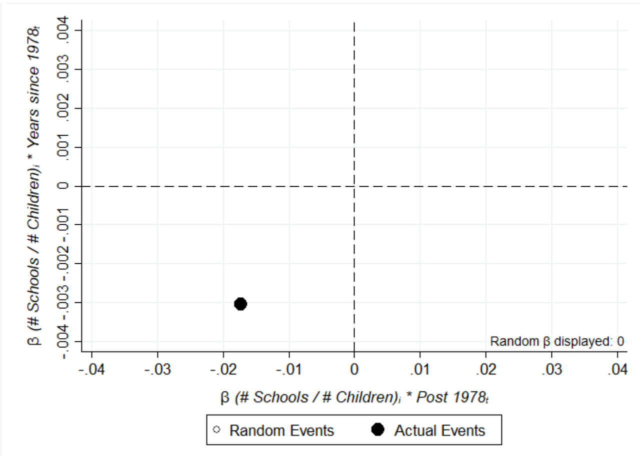
A) All estimated coefficients



B) Both coefficients stat. sign. at the 10 % level



C) Both coefficients stat. sign. at the 5 % level



C) Both coefficients stat. sign. at the 1 % level

Note - Each panel displays all coefficients obtained using 1,000 placebo conflict datasets with the same average conflict likelihood as our main conflict dataset built based on SMH articles. Each dot corresponds to one combination of coefficients in a cartesian plane where the horizontal axis represents the beta coefficient of the specification of Column 3, while the vertical axis depicts the beta coefficient of the specification of Column 6 of baseline Table 2. The large black dot represents our true coefficients. Panel A displays all coefficients. Panel B (C) [D] shows the estimates when the coefficients obtained with the two specifications (and the same placebo dataset) are both statistically significant at the 10 % (5 %) [1 %] level. The number of placebo dataset displayed in each cartesian plan is reported in the bottom-right corner.

B.10 Robustness Analysis: Reclink Match Locations

Below we replicate our main results using an alternative reclink threshold. Specifically, we assess whether our results hold when alternative matching scores are adopted in the matching of entities and locations in the gazetteer. Tables B9 and B10 display the results. Estimates reported in the first columns correspond to our baseline estimates obtained using a matching score equal to one (i.e., a perfect match between entities and gazetteer' locations). The remaining columns show results obtained using a fuzzy match and are ranked based on the distance from the perfect matching score. Reassuringly the results are qualitatively similar to our preferred estimates. The pattern of somewhat smaller magnitudes obtained using less accurate matching scores is consistent with the view that noisy coverage of conflict events may lead to an attenuation bias.

Table B9: Alternative reclink score threshold in the geographical matching 1/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0127***	-0.0123***	-0.0129***	-0.00961**	-0.00923*	-0.0105**	-0.0113**	-0.0115**	-0.0113**	-0.0130**	-0.0132**
	(0.00448)	(0.00446)	(0.00460)	(0.00485)	(0.00486)	(0.00504)	(0.00509)	(0.00508)	(0.00510)	(0.00520)	(0.00520)
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reclink Score	1.000	0.998	0.996	0.994	0.992	0.990	0.988	0.986	0.984	0.982	0.980
Sample Mean	.0773	.0785	.0831	.0927	.097	.1131	.117	.1173	.1275	.1411	.1414

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. Estimates reported in the first column correspond to our baseline estimates obtained using a matching score equal to one (i.e., a perfect match between entities and gazetteer' locations). The remaining columns show results obtained using a fuzzy match and are ranked based on the distance from the perfect matching score. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B10: Alternative reclink score threshold in the geographical matching 2/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
*(# Schools / # Children)$_i$ * Years since 1978$_t$*	-0.00146***	-0.00143***	-0.00152***	-0.00132***	-0.00130***	-0.00154***	-0.00164***	-0.00167***	-0.00163***	-0.00179***	-0.00182***
	(0.000421)	(0.000419)	(0.000435)	(0.000440)	(0.000440)	(0.000459)	(0.000464)	(0.000463)	(0.000465)	(0.000470)	(0.000470)
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reclink Score	1.000	0.998	0.996	0.994	0.992	0.990	0.988	0.986	0.984	0.982	0.980
Sample Mean	.0773	.0785	.0831	.0927	.097	.1131	.117	.1173	.1275	.1411	.1414

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. Estimates reported in the first column correspond to our baseline estimates obtained using a matching score equal to one (i.e., a perfect match between entities and gazetteer' locations). The remaining columns show results obtained using a fuzzy match and are ranked based on the distance from the perfect matching score. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.11 Robustness Analysis: Alternative Location Coding

Below we present a replication of the main baseline results when using an alternative location coding mechanism. In particular, we use the *geotex* module to identify locations. The *geotex* module is "faster" then the *NLTK* module but is less accurate mainly because it relies on a pre-defined library with a list of places that is not extensive. Table B11 reports estimates obtained using this alternative algorithm to identify locations. Due to the low accuracy, the average incidence of conflict events obtained with this second routine is smaller than with our preferred algorithm. However, we find that our findings are overall robust to this sensitivity check.

Table B11: Robustness: Alternative algorithm adopted to code locations

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.00570*	0.000826	-0.00191						
	(0.00310)	(0.00268)	(0.00370)						
*(# Schools / # Children)$_i$ * Years since 1978$_t$*				-0.000805**	-0.000814*	-0.00120**			
				(0.000311)	(0.000474)	(0.000543)			
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$*							-0.000241	0.00146	-0.00162
							(0.00367)	(0.00287)	(0.00422)
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$*							-0.00540*	-0.00307	-0.00581
							(0.00324)	(0.00380)	(0.00415)
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$*							-0.0125***	-0.00963	-0.0171**
							(0.00471)	(0.00649)	(0.00733)
Observations	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560	11,560
R-squared	0.371	0.450	0.494	0.372	0.451	0.494	0.372	0.451	0.495
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Province x Year FEs	No	No	Yes	No	No	Yes	No	No	Yes
Sample Mean	.03	.03	.03	.03	.03	.03	.03	.03	.03

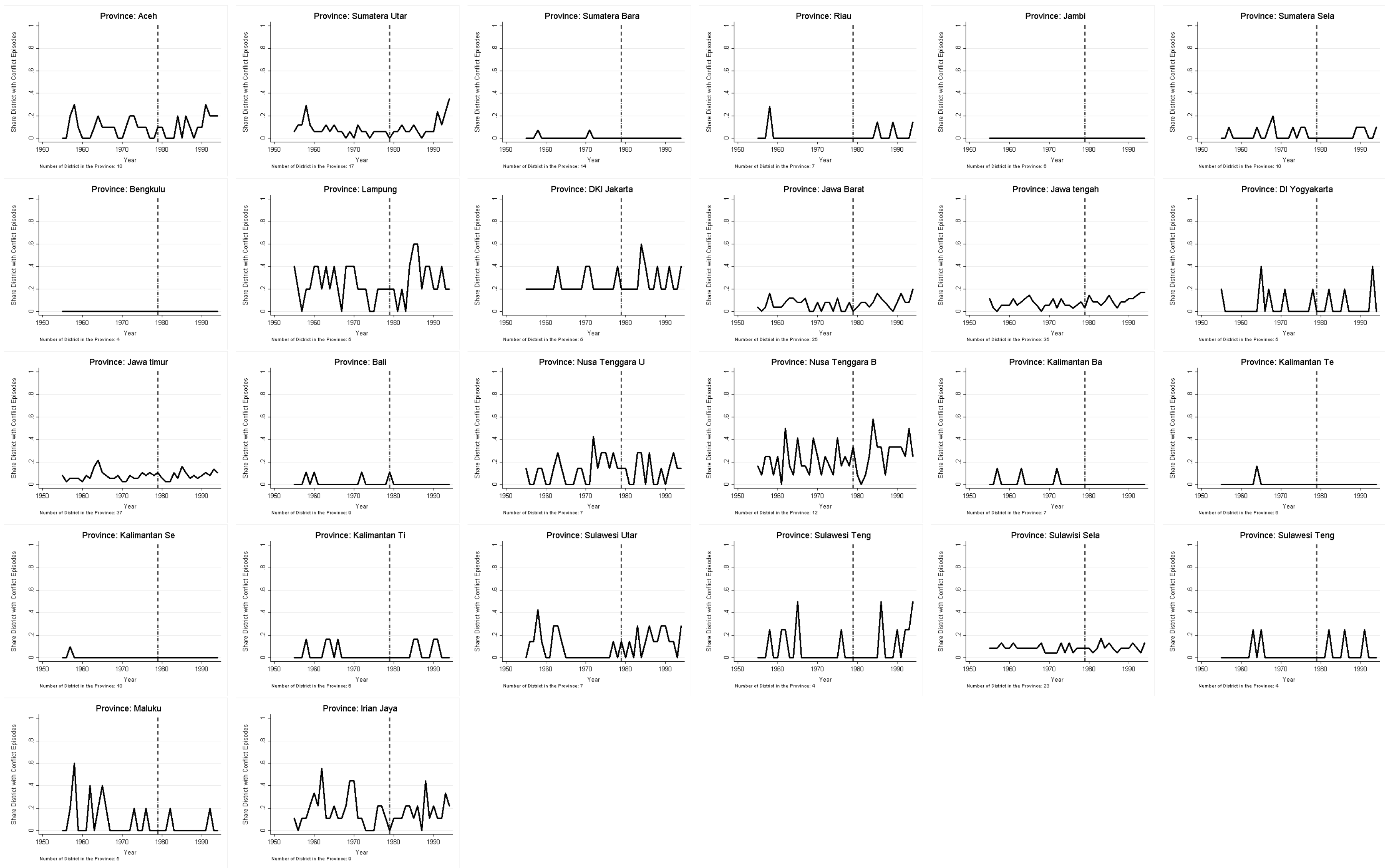
NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Years 1979-1984$_t$* is a dummy taking a value of 1 for the years 1979-1984 (it is analogous for the two variables referring to the period 1985-1989 and 1990-1994, respectively). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and using an alternative location coding mechanism. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.12 Robustness Analysis: Outliers and sample composition

As discussed in Section 6.9, we display below a series of robustness results when removing data from one province at a time or cropping our sample duration. We start by describing graphically in Figure B3 the evolution of the share of districts per province and year that have experienced conflict. It turns out that for many provinces conflict is stationary over time and does not show any particular trend. Interestingly, in some provinces such as Aceh and its neighboring province Sumatera Utar there is an uptake in conflict events around the early 1990s when the Indonesian government stepped up repression of the Aceh independence movement.

In Table B12 we replicate our baseline regressions when dropping one province at a time, while in Tables B13 and B14 we similarly investigate the robustness of our baseline findings to reducing the length of our sample duration (to address potential concerns about the surge in conflict at the beginning of the 1990s). Reassuringly, the results are hardly changed in any of these sensitivity checks.

Figure B3: Share of districts within each province with conflict events over time



SOURCE: Authors' computations from own conflict data. Conflict data is obtained using the procedure described in Section 4.1.

Table B12: Robustness: Dropping one province at the time

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.00305***	-0.00296***	-0.00234***	-0.00314***	-0.00298***	-0.00308***	-0.00308***	-0.00305***	-0.00311***	-0.00304***	-0.00302***	-0.00310***	-0.00305***	-0.00318***
	(0.000705)	(0.000716)	(0.000759)	(0.000727)	(0.000703)	(0.000712)	(0.000708)	(0.000705)	(0.000706)	(0.000704)	(0.000720)	(0.000714)	(0.000705)	(0.000719)
Observations	11,560	11,160	10,880	11,000	11,280	11,320	11,160	11,400	11,360	11,360	10,560	10,160	11,360	10,080
R-squared	0.506	0.514	0.496	0.507	0.508	0.506	0.510	0.506	0.499	0.489	0.517	0.511	0.508	0.511
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Drop Province	-	Aceh	Sumat Utar	Sumat Bara	Riau	Jambi	Sumat Sela	Bengkulu	Lampung	DKI Jakarta	Jawa Barat	Jawa Tengah	DI Yogyak	Jawa Timur
Sample Mean	.077	.077	.077	.081	.079	.079	.079	.078	.074	.074	.078	.077	.078	.077

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.00305***	-0.00343***	-0.00307***	-0.00301***	-0.00301***	-0.00322***	-0.00308***	-0.00319***	-0.00280***	-0.00302***	-0.00308***	-0.00305***	-0.00314***	-0.00297***
	(0.000705)	(0.000755)	(0.000708)	(0.000704)	(0.000708)	(0.000745)	(0.000714)	(0.000762)	(0.000698)	(0.000704)	(0.000739)	(0.000707)	(0.000749)	(0.000710)
Observations	11,560	11,200	11,280	11,080	11,280	11,320	11,160	11,320	11,280	11,400	10,640	11,400	11,360	11,200
R-squared	0.506	0.508	0.509	0.505	0.508	0.507	0.507	0.508	0.511	0.508	0.496	0.508	0.508	0.501
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Drop Province	-	Bali	Nusa Teng T	Nusa Teng B	Kali Ba	Kali Te	Kali Se	Kali Ti	Sul Utar	Sul Teng	Sul Sela	Sul Yeng	Maluku	Irian Jaya
Sample Mean	.077	.079	.076	.071	.079	.079	.08	.078	.077	.077	.076	.078	.077	.074

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. Estimates obtained with the full sample are reported in the first columns. Remaining columns report estimates obtained by dropping one province at time. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B13: Robustness: Shortening the period of interest by one year at the time 1/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0127***	-0.0173***	-0.0113**	-0.0188***	-0.0104**	-0.0193***	-0.0101**	-0.0197***	-0.00915**	-0.0199***	-0.00862*	-0.0198***
	(0.00448)	(0.00610)	(0.00444)	(0.00616)	(0.00444)	(0.00618)	(0.00443)	(0.00621)	(0.00441)	(0.00621)	(0.00441)	(0.00620)
Observations	11,560	11,560	11,271	11,271	10,982	10,982	10,693	10,693	10,404	10,404	10,115	10,115
R-squared	0.345	0.506	0.347	0.506	0.347	0.504	0.348	0.506	0.347	0.506	0.347	0.506
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Sample Period	1955-1994	1955-1994	1955-1993	1955-1993	1955-1992	1955-1992	1955-1991	1955-1991	1955-1990	1955-1990	1955-1989	1955-1989
Sample Mean	.077	.077	.076	.076	.074	.074	.074	.074	.073	.073	.072	.072

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. Estimates obtained with the full sample are reported in the first column. Remaining columns report estimates obtained by shortening the period of interest by one year at the time. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B14: Robustness: Shortening the period of interest by one year at the time 2/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*	-0.00146***	-0.00305***	-0.00131***	-0.00271***	-0.00122***	-0.00267***	-0.00130**	-0.00254***	-0.00120**	-0.00222***	-0.00120**	-0.00219***
	(0.000421)	(0.000705)	(0.000442)	(0.000707)	(0.000471)	(0.000730)	(0.000504)	(0.000749)	(0.000539)	(0.000787)	(0.000590)	(0.000807)
Observations	11,560	11,560	11,271	11,271	10,982	10,982	10,693	10,693	10,404	10,404	10,115	10,115
R-squared	0.346	0.506	0.347	0.506	0.347	0.504	0.348	0.506	0.347	0.506	0.347	0.506
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Trend	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Province x Year FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Sample Period	1955-1994	1955-1994	1955-1993	1955-1993	1955-1992	1955-1992	1955-1991	1955-1991	1955-1990	1955-1990	1955-1989	1955-1989
Sample Mean	.077	.077	.076	.076	.074	.074	.074	.074	.073	.073	.072	.072

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. Estimates obtained with the full sample are reported in the first column. Remaining columns report estimates obtained by shortening the period of interest by one year at the time. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.13 Heterogeneous Effects: Type of Conflict Keywords

Below are listed the keywords that are used in the analysis of heterogeneous effects with respect to different types of conflict carried out in Section 7. In particular, to construct the "economic", "religious/ethnic" and "political" conflict variables, we proceed as follows.

We used natural language processing algorithms to analyze the content of all articles containing conflict-related events (see additional details in Section 4.1 and in Online Appendix B.1). In doing so, we searched in the text for a three different set of terms used to identify "economic", "religious/ethnic" or "political" conflict events (reported in Tables B15, B16 and B17, respectively). Categories are not exclusive and an event can be related to more than one type.

Table B15: Keywords used for constructing economic conflict

Economic, Job, Unemployment, Recession, Income, Wage, Salary, Growth, Industry, Food, Price, Famine, Starvation, Scarcity, Poverty.

Table B16: Keywords used for constructing ethno-religious conflict

Muslim, Protestant, Catholic, Hindu, Buddhist, Temple, Church, Mosque, Candi, Masjid, Religion, Religious, Faith, Sundanese, Malay, Madurese, Batak, Minangkabau, Sundanese, Malay, Madurese, Batak, Minangkabau, Betawi, Bugis, Acehnese, Bantenese, Banjarese, Balinese, Chinese, Sasak, Makassarese, Minahasan, Cirebonese, Ethnicity, Ethnic, Tribe, Tribal, Linguist, Language, Identity, Cultural, Tradition.

Table B17: Keywords used for constructing political conflict

Election, Vote, Mayor, Government, Corruption, Bribery, Politics.

B.14 Heterogeneous Effects: Additional heterogeneous effects on rural population and on bride price

Below in Table B18 we present further heterogeneous effect estimations with respect to the level of rural population and the practice of bride prices. We find that there are no discernable heterogeneous effects and that our results hold in both rural and urban areas, as well as both in the presence and in the absence of bride prices.

Table B18: Heterogeneous effects on rural population and on bride price

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0181**		-0.00916**	
	(0.00710)		(0.00459)	
*(# Schools / # Children)$_i$ * Years Since 1978$_t$*		-0.00455***		-0.00249***
		(0.00110)		(0.000737)
*(# Schools / # Children)$_i$ * Post-1978$_t$ * High Intensity of Rural Population$_i$*	-7.25e-05			
	(0.00986)			
*(# Schools / # Children)$_i$ * Years Since 1978$_t$ * High Intensity of Rural Population$_i$*		0.00209		
		(0.00135)		
*(# Schools / # Children)$_i$ * Post-1978$_t$ * High Intensity of the Practice of Bride Price$_p$*			-0.0105	
			(0.0140)	
*(# Schools / # Children)$_i$ * Years Since 1978$_t$ * High Intensity of the Practice of Bride Price$_p$*				-2.61e-07
				(0.00207)
Observations	11,560	11,560	10,800	10,800
R-squared	0.506	0.507	0.511	0.511
District FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
District-Specific Linear Trend	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes
Sample Mean	.08	.08	.07	.07

NOTE: The unit of observation is a district i and year t. The full sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The unit of observation is a district i. The variable *High Intensity of Rural Population$_i$* is a dummy that takes a value of 1 if more than half of individuals in district i live in rural areas. The share of the population of district i living in rural areas was computed using the 1971 Census (IPUMS (2018)). The variable *High Intensity of the Practice of Bride Price$_p$* is a dummy that takes a value of 1 if more than 50 percent of individuals in province p are from an ethnic group that traditionally practices bride price as opposed to other customs (Source: Ashraf et al. (2019)). The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Columns (1) and (2) include the variables (*High Intensity of Rural Population$_i$ * Post-1978$_t$*) and (*High Intensity of Rural Population$_i$ * Years Since 1978$_t$*), respectively. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.15 Channels and Mechanisms: Economic returns versus religious cleavages

In this Online Appendix we produce a series of sensitivity checks for the analysis carried out in the main text in Section 8.1. In particular, we present a replication of the main results when economic and societal variables are computed i) dropping all individuals attending school and ii) using the full sample from the Indonesian population census of 1971 (IPUMS (2018)). Reassuringly the results are similar to our preferred estimates (see Table B19 below).

We also show robustness to the major robustness checks carried out earlier on in the Appendix A.8, namely the broadening of the keywords used (see Table B20 below) as well as the inclusion of the *Canberra Times* as alternative media source (see Table B21 below). In both cases, our results remain very similar. Finally, we also report below Table B22 where we slice the treatment period in three subperiods in the aim of documenting the evolution of the effects over time, finding that the economic returns to education channel gains importance over time. The coefficients of this table have been represented graphically in Figure 3 in the main text.

Table B19: Robustness mechanism: Alternative sample

	Drop individuals in school				Full sample			
Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(# Schools / # Children)$_i$ * *Post-1978*$_t$	-0.00844	-0.00359			-0.0103	-0.00796		
	(0.00651)	(0.00610)			(0.00629)	(0.00594)		
(# Schools / # Children)$_i$ * *Years since 1978*$_t$			-0.00258***	-0.00214**			-0.00281***	-0.00271***
			(0.000134)	(0.0000843)			(0.000783)	(0.000819)
(# Schools / # Children)$_i$ * *Post-1978*$_t$ * *Religious Polarization*$_i$	-0.0503**	-0.0416*			-0.0444**	-0.0292		
	(0.0222)	(0.0233)			(0.0210)	(0.0181)		
(# Schools / # Children)$_i$ * *Post-1978*$_t$ * *Average RoE Bricks and Entrep.*$_i$	-0.00696				0.000423			
	(0.0100)				(0.00706)			
(# Schools / # Children)$_i$ * *Post-1978*$_t$ * *Princ. Comp. RoE Bricks and Entrep.*$_i$		-0.00806				0.00276		
		(0.00904)				(0.00651)		
(# Schools / # Children)$_i$ * *Years since 1978*$_t$ * *Religious Polarization*$_i$			-0.00800**	-0.00945***			-0.00722***	-0.00656***
			(0.00119)	(0.00323)			(0.00234)	(0.00246)
(# Schools / # Children)$_i$ * *Years since 1978*$_t$ * * *Average RoE Bricks and Entrep.*$_i$			-0.00274**				-0.00274**	
			(0.00120)				(0.00120)	
(# Schools / # Children)$_i$ * *Years since 1978*$_t$ * *Princ. Comp. RoE Bricks and Entrep.*$_i$				-0.00203*				0.000897
				(0.00109)				(0.000847)
Observations	10,600	8,240	10,600	8,240	10,600	8,800	10,600	8,800
R-squared	0.521	0.540	0.522	0.542	0.521	0.539	0.522	0.540
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.81	.080	.081	.080	.082	.081	.081	.083

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)*$_i$ represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978*$_t$ takes a value of 1 for the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable defined as *Years since 1978*$_t$ is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Religious Polarization* corresponds to the level of religious polarization in district i, whereas the variable *Return to Education* (*RoE*) indicates the relative economic advantages at the district level from having completed primary school. Religious polarization and returns to education measures were computed using the 1971 Census (IPUMS (2018)) (see additional details in the text). Columns 1-4 (5-9) display results obtained when economic and societal variables are computed dropping all individuals attending school (using the full sample). The conflict data was constructed using the *Sydney Morning Herald* and the *Canberra Times* (i.e. a district year was coded as having a conflict when there was a corresponding article in any of these two newspapers), following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B20: Robustness mechanism: Alternative sources 1/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	0.00535	0.0114	0.00527	0.0139				
	(0.0102)	(0.00832)	(0.00934)	(0.00904)				
*(# Schools / # Children)$_i$ * Years since 1978$_t$*					-0.00213**	-0.00172*	-0.00214**	-0.00166
					(0.00103)	(0.000954)	(0.000934)	(0.00102)
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Religious Polarization$_i$*	-0.0427	-0.0249	-0.0424	-0.0438				
	(0.0290)	(0.0244)	(0.0291)	(0.0305)				
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Return to Education [Bricks]$_i$*	-0.0439							
	(0.0268)							
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Return to Education [Entrep.]$_i$*		-0.00245						
		(0.00731)						
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Average RoE Bricks and Entrep.$_i$*			-0.00975					
			(0.0145)					
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*				-0.0206				
				(0.0137)				
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Religious Polarization$_i$*					-0.00890***	-0.00565**	-0.00813***	-0.00825***
					(0.00318)	(0.00251)	(0.00295)	(0.00282)
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Return to Education [Bricks]$_i$*					-0.00610***			
					(0.00231)			
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Return to Education [Entrep.]$_i$*						-0.000989		
						(0.000620)		
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * * Average RoE Bricks and Entrep.$_i$*							-0.00241*	
							(0.00126)	
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*								-0.00317**
								(0.00134)
Observations	9,480	9,040	10,480	8,040	9,480	9,040	10,480	8,040
R-squared	0.554	0.553	0.552	0.555	0.556	0.553	0.553	0.556
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.135	.126	.132	.129	.135	.126	.132	.129

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Religious Polarization* corresponds to the level of religious polarization in district i, whereas the variable *Return to Education* (*RoE*) indicates the relative economic advantages at the district level from having completed primary school. Religious polarization and returns to education measures were computed using the 1971 Census (IPUMS (2018)) (see additional details in the text). Estimates presented in columns 1-4 (5-8) are obtained by removing from the Census all individuals attending school (by using the full sample). The conflict data was constructed using the *Sydney Morning Herald* and a broader set of conflict-related keywords (see discussion in Appendix A.7). Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B21: Robustness mechanism: Alternative sources 2/2

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*(# Schools / # Children)$_i$ * Post-1978$_t$*	-0.0136*	-0.00983	-0.0119	-0.00982				
	(0.00771)	(0.00748)	(0.00747)	(0.00723)				
*(# Schools / # Children)$_i$ * Years since 1978$_t$*					-0.00323***	-0.00333***	-0.00313***	-0.00348***
					(0.000949)	(0.000889)	(0.000849)	(0.00100)
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Religious Polarization$_i$*	-0.0271	-0.0194	-0.0317	-0.0204				
	(0.0254)	(0.0240)	(0.0267)	(0.0261)				
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Return to Education [Bricks]$_i$*	-0.0110							
	(0.0207)							
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Return to Education [Entrep.]$_i$*		0.00440						
		(0.00689)						
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Average RoE Bricks and Entrep.$_i$*			0.00372					
			(0.0125)					
*(# Schools / # Children)$_i$ * Post-1978$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*				-0.00178				
				(0.0135)				
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Religious Polarization$_i$*					-0.00828***	-0.00691***	-0.00853***	-0.00845***
					(0.00295)	(0.00239)	(0.00248)	(0.00267)
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Return to Education [Bricks]$_i$*					-0.00380*			
					(0.00205)			
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Return to Education [Entrep.]$_i$*						-0.00123**		
						(0.000559)		
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * * Average RoE Bricks and Entrep.$_i$*							-0.00230**	
							(0.00105)	
*(# Schools / # Children)$_i$ * Years since 1978$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*								-0.00287***
								(0.00107)
Observations	9,480	9,040	10,480	8,040	9,480	9,040	10,480	8,040
R-squared	0.556	0.561	0.553	0.563	0.557	0.562	0.554	0.565
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Mean	.11	.102	.107	.105	.11	.102	.107	.105

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The dummy *Post-1978$_t$* takes a value of 1 for the years after the first year when we expect the program to deploy major effects (which is when the first INPRES cohort reaches the critical age for being recruitable for fighting – see discussion in Section 5.3). The variable defined as *Years since 1978$_t$* is a variable that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. The variable *Religious Polarization* corresponds to the level of religious polarization in district i, whereas the variable *Return to Education* (*RoE*) indicates the relative economic advantages at the district level from having completed primary school. Religious polarization and returns to education measures were computed using the 1971 Census (IPUMS (2018)) (see additional details in the text). The conflict data was constructed using the *Sydney Morning Herald* and the *Canberra Times* (i.e. a district year was coded as having a conflict when there was a corresponding article in any of these two newspapers), following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B22: Robustness mechanism: Effect by sample subperiods

Dep. Variable: Conflict Episode$_{it}$	(1)	(2)	(3)	(4)
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$*	-0.0142	0.00181	-0.00280	0.00161
	(0.0100)	(0.00751)	(0.00889)	(0.00654)
*(# Schools / # Children)$_i$ * Years 1985-1989$_{tt}$*	-0.0283*	-0.00931	-0.0147	-0.00812
	(0.0158)	(0.0121)	(0.0142)	(0.0133)
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$*	-0.0459**	-0.0293**	-0.0427**	-0.0135
	(0.0221)	(0.0138)	(0.0186)	(0.0149)
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$ * Religious Polarization$_i$*	-0.0399**	-0.0398**	-0.0490***	-0.0407**
	(0.0192)	(0.0168)	(0.0186)	(0.0170)
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$ * Religious Polarization$_i$*	-0.0494	-0.0367	-0.0561*	-0.0447
	(0.0331)	(0.0294)	(0.0322)	(0.0330)
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$ * Religious Polarization$_i$*	-0.0984**	-0.108***	-0.114***	-0.117***
	(0.0387)	(0.0323)	(0.0331)	(0.0346)
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$ * Return to Education [Bricks]$_i$*	-0.0209**			
	(0.0105)			
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$ * Return to Education [Bricks]$_i$*	-0.0346*			
	(0.0200)			
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$ * Return to Education [Bricks]$_i$*	-0.0398**			
	(0.0191)			
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$ * Return to Education [Entrep.]$_i$*		-0.00193		
		(0.00442)		
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$ * Return to Education [Entrep.]$_i$*		-0.00185		
		(0.00629)		
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$ * Return to Education [Entrep.]$_i$*		-0.0159**		
		(0.00720)		
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$ * Average RoE Bricks and Entrep.$_i$*			-0.00643	
			(0.00785)	
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$ * Average RoE Bricks and Entrep.$_i$*			-0.0112	
			(0.0125)	
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$ * Average RoE Bricks and Entrep.$_i$*			-0.0318**	
			(0.0126)	
*(# Schools / # Children)$_i$ * Years 1979-1984$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*				-0.0111
				(0.00682)
*(# Schools / # Children)$_i$ * Years 1985-1989$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*				-0.0169
				(0.0135)
*(# Schools / # Children)$_i$ * Years 1990-1994$_t$ * Princ. Comp. RoE Bricks and Entrep.$_i$*				-0.0317***
				(0.0101)
Observations	9,480	9,040	10,480	8,040
R-squared	0.524	0.530	0.519	0.536
District FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
District-Specific Linear Time Trend	Yes	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes	Yes

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1955-1994. LPM estimates are reported in all columns. The dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years 1979-1984$_t$* is a dummy taking a value of 1 for the years 1979-1984 (it is analogous for the two variables referring to the period 1985-1989 and 1990-1994, respectively). The variable *Religious Polarization* corresponds to the level of religious polarization in district i, whereas the variable *Return to Education* (*RoE*) indicates the relative economic advantages at the district level from having completed primary school. Religious polarization and returns to education measures were computed using the 1971 Census (IPUMS (2018)) (see additional details in the text). The conflict data was constructed using the *Sydney Morning Herald*, following the approach described in Section 4.1 and in Online Appendix B.1. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.16 Channels and Mechanisms: How education may attenuate religious tensions

In this Online Appendix we carry out sensitivity checks for the estimations of Section 8.2. In particular, in Table B23 we replicate the main analysis but code the dependent variables as dummies instead of continuous variables, while in Table B24 we adopt the same sample as in Duflo (2001) (covering only the years 1950 to 1972 instead of 1945 to 1972, which is the sample period in our main analysis of Section 8.2).

Finally, we also interact in Table B25 our exposure to education variable with being of Muslim or Christian religion (with all other religious denominations being the omitted category) to see whether education may have a bigger or smaller impact for people belonging to the dominant religious denomination of the country (i.e. about 87 percent of Indonesia's population are Muslim and 10 percent Christians). We detect no differential effects and find that education breeds tolerance to a similar extent for different religions.

Table B23: Robustness - Societal channels: Alternative coding of survey answers

Dep. Variable:	(1) *Trust Dummy$_n$*	(2) *Trust Dummy$_n$*	(3) *Marriage Dummy$_n$*	(4) *Marriage Dummy$_n$*	(5) *Relig. Dummy$_n$*	(6) *Relig. Dummy$_n$*
*(# Schools / # Children)$_i$ * Born after 1962$_n$*	0.0173**	0.0158**	0.0121**	0.00819	0.0110*	0.00963
	(0.00666)	(0.00676)	(0.00555)	(0.00541)	(0.00637)	(0.00704)
Observations	10,573	10,573	10,574	10,574	10,547	10,547
R-squared	0.126	0.159	0.190	0.250	0.082	0.094
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	Yes	No	Yes	No	Yes
Ethnicity	No	Yes	No	Yes	No	Yes
Religion	No	Yes	No	Yes	No	Yes

NOTE: The unit of observation is an individual n born in district i. The sample covers all individuals surveyed in the *Wave 5* of the IFLS SURVEY, born between 1945 and 1972. LMP estimates are reported in all columns. The dependent variables correspond to the dichotomous version of variables *Trust*, *Marriage* and *Religiosity* used in Table 7 and 8, with values 0-1 being coded as 0, and values 2-3 coded as 1. Additional details on survey variables are provided in 8.2. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Born after 1962$_n$* is a dummy that takes a value of 1 if a given individual n was born after 1962 in district i. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B24: Robustness - Societal channels: Alternative sample

Dep. Variable:	(1) $Trust_n$	(2) $Marriage_n$	(3) $Roscas_n$	(4) $Trust_n$	(5) $Marriage_n$	(6) $Roscas_n$
*(# Schools / # Children)$_i$ * Born after 1962$_n$*	0.0342***	0.0322**	0.0175**	0.0320**	0.0280**	0.0162**
	(0.0127)	(0.0145)	(0.00739)	(0.0131)	(0.0133)	(0.00652)
Observations	9,837	9,838	10,434	9,837	9,838	9,787
R-squared	0.125	0.187	0.153	0.151	0.230	0.239
District FEs	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	No	No	Yes	Yes	Yes
Ethnicity	No	No	No	Yes	Yes	Yes
Religion	No	No	No	Yes	Yes	Yes

NOTE: The unit of observation is an individual n born in district i. The sample covers all individuals surveyed in the *Wave 5* of the IFLS SURVEY, born between 1950 and 1972. OLS estimates are reported in all columns. $Trust_n$ and $Marriage_n$ variables are used as continuous variables ranging from 0 to 3, treating the scales of the survey questions as cardinal. $Roscas_n$ is a dummy that take a value of 1 if the individual participated to a *arisan* community group over the previous 12 months. Additional details on survey variables are provided in the text. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Born after 1962$_n$* is a dummy that takes a value of 1 if a given individual n was born after 1962 in district i. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B25: Heterogeneous effect - Societal channels: Type of religions

Dep. Variable:	(1) $Trust_n$	(2) $Marriage_n$	(3) $Arisan_n$
*(# Schools / # Children)$_i$ * Born after 1962$_n$*	0.0393	0.0282**	0.0232***
	(0.0259)	(0.0123)	(0.00804)
*(# Schools / # Children)$_i$ * Born after 1962$_n$ * Muslim$_n$*	-0.0154	-0.00111	-0.0139
	(0.0277)	(0.0141)	(0.00938)
*(# Schools / # Children)$_i$ * Born after 1962$_n$ * Christian$_n$*	0.00683	-0.0121	0.00252
	(0.0330)	(0.0207)	(0.0179)
Observations	10,576	10,577	10,516
R-squared	0.133	0.226	0.237
District FEs	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes
Gender	Yes	Yes	Yes
Ethnicity	Yes	Yes	Yes
Religion	Yes	Yes	Yes

NOTE: The unit of observation is an individual n born in district i. The sample covers all individuals surveyed in the *Wave 5* of the IFLS SURVEY, born between 1945 and 1972. OLS estimates are reported in all columns. $Trust_n$ and $Marriage_n$ variables are used as continuous variables ranging from 0 to 3, treating the scales of the survey questions as cardinal. $Roscas_n$ is a dummy that take a value of 1 if the individual participated to a *arisan* community group over the previous 12 months. Additional details on survey variables are provided in the text. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Born after 1962$_n$* is a dummy that takes a value of 1 if a given individual n was born after 1962 in district i. The variable *Muslim$_n$* is a dummy that takes a value of 1 if a given individual n is affiliated with the Muslim religion. The variable *Christian$_n$* is a dummy that takes a value of 1 if a given individual n is affiliated with a Christian religion. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.17 Channels and Mechanisms: How education may attenuate ethnic tolerance

Below in Table B26 we replicate the analysis of the impact of schooling on religious tolerance, but now focusing on another question of the survey referring to ethnic instead of religious tolerance. In particular, we make use of the question "Taking into account the diversity of ethnicities in the village, I trust people withe same ethnicity as mine more", where again the answer option range from 1 (Strongly Agree) to 4 (Strongly disagree).

We find that ethnic tolerance is not affected by schooling. One possible explanation could be that the Indonesian state ideology *Pancasila* – which was taught in school – stresses as one of the five fundamental principles the importance of religious freedom and tolerance, while an analogous emphasis on ethnic relations is absent. Another plausible explanation is that traditionally inter-ethnic trust has been significantly higher in Indonesia than inter-religious trust (which may have triggered the particular emphasis of *Pancasila* on addressing the main source of tension: religious sectarianism). Indeed, in our IFLS survey wave only 18.6 percent of respondents are classified as having high levels of inter-religious trust, while for inter-ethnic trust the number is much larger (32.4 percent).

Table B26: Societal channels: Ethnic tolerance

Dep. Variable: Trust - Ethnicity$_n$	(1)	(2)	(3)	(4)
*(# Schools / # Children)$_i$ * Post-1962 Cohorts$_n$*	-0.00483	-0.00281	-0.00550	-0.00376
	(0.0159)	(0.0160)	(0.0165)	(0.0166)
Observations	10,531	10,531	9,790	9,790
R-squared	0.095	0.104	0.095	0.104
District FEs	Yes	Yes	Yes	Yes
Cohort FEs	Yes	Yes	Yes	Yes
Gender	No	Yes	No	Yes
Ethnicity	No	Yes	No	Yes
Religion	No	Yes	No	Yes
Cohorts	1945-1973	1945-1973	1950-1973	1950-1973

NOTE: The unit of observation is an individual n born in district i. The sample covers all individuals surveyed in the *Wave 5* of the IFLS SURVEY. Estimates reported in the first (last) two columns are obtained using all individuals born between 1945 and 1972 (1950 and 1972). OLS estimates are reported in all columns. *Trust - Ethnicity$_n$* is used as continuous variables ranging from 0 to 3, treating the scales of the survey questions as cardinal. Additional details are provided in the text. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Born after 1962$_n$* is a dummy that takes a value of 1 if a given individual n was born after 1962 in district i. Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.18 Channels and Mechanisms: Voice versus violence

In this Online Appendix we discuss the construction of a measure of non-violent events using the Sydney Morning Herald. Our approach to construct a geo-referenced dataset of non-violent related events in Indonesia is very similar to the methodology used to identify conflict-related events in the main analysis. The only difference lies in the set of keywords, as to identify non-violent related events we use the keywords "demonstration", "march", "gather", "manifestation" and "picket". We used natural language processing algorithms to analyse the content of all articles, storing all sentences with at least one non-violent, related term (excluding those where a conflict related term was also present). Finally, we started out using a Named Entity Recognition algorithm to identify all locations referred to, and then matched locations to geo-coordinates.

Reassuringly, our measure yields very similar values as the established dataset GDELT for the years of temporal overlap (1979-1994). The protest measure of GDELT takes a mean value of 0.021 while our measure has an average of 0.020. Importantly, in around 95 percent of cases our non-violent episodes variable takes the same value as the GDELT protest measure (both 0 or both 1).

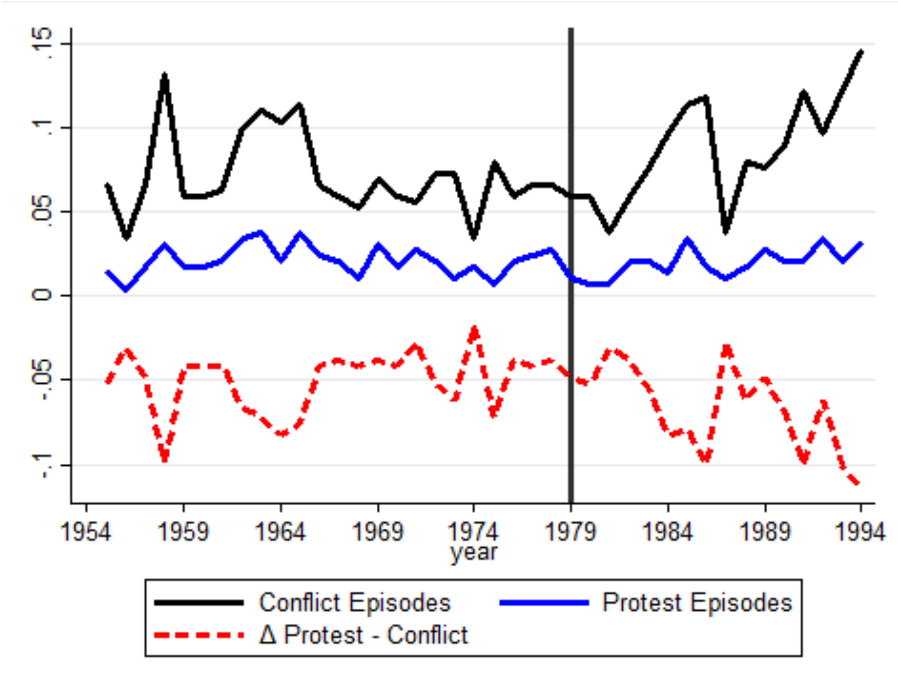
The specification we run is given by

$$Dep.\ Variable_{it} = \alpha + \beta \frac{\#Schools\ Built}{\#Children}_i * Years\ since\ 1978_t + FE_i + FE_t + \epsilon_{it},$$

with the three different dependent variables (defined at the district-year level) being the following: "Conflict Episode" (which corresponds to the dependent variables used in the main analysis), "Pacific Episode" (coded using the aforementioned keywords), and finally "Δ Pacific Episode - Conflict Episode" (coded as the difference between the two variables defined above). This last, *relative* measure can take values 1 (in a district-year with pacific events and no conflict events), 0 (in a district-year with both pacific and conflict events or no pacific and no conflict events), or -1 (with conflict events and no pacific events). All other variables are defined as above in the baseline analysis. Figure B4 plots the evolution over time of these measures.

Table 9 in the main text has displayed the main results on "violence versus voice". Below we display a set of additional tables, replicating among others the results for the alternative datasets GDELT and ICEWS, in Tables B27 and B28, respectively. While for GDELT data we find a quantitatively small protest-decreasing effect for education, with ICEWS we find the same non-result on protests as with our main measure. Importantly, in all cases we always find that Δ *Pacific - Conflict*, the relative scope for peaceful protests rather than violent conflict, is increased by more INPRES schools (imprecisely estimated for GDELT, and statistically significant for

Figure B4: Evolution of conflict and protest events over time



SOURCE: Authors' computations from own conflict and (pacific) protest data. Conflict data is obtained using the procedure described in Section 4.1. Protest data is obtained using the procedure described in Section 8.3.

ICEWS).

Table B27: Conflict events vs protest events: Alternative sources (GDELT (2018) data)

	(1)	(2)	(3)
Dep. Variable:	*Conflict Episode$_{it}$*	*Protest Episode$_{it}$*	Δ *Protest - Conflict$_{it}$*
*(# Schools / # Children)$_i$ * Years since 1978$_t$*	-0.00201***	-0.00147***	0.000537
	(0.000628)	(0.000452)	(0.000674)
Observations	10,404	10,404	10,404
R-squared	0.624	0.426	0.374
District FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes
Sample Mean	.41	.11	-.3

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1979-2014. LPM estimates are reported in all columns. In column 1 the dependent variable is a dummy that takes a value of 1 if a violent conflict event was observed in district i and year t. Column 2 has as dependent variable a dummy that takes a value of 1 if a peaceful protest was observed in district i and year t. In column 3 the dependent variable is the difference between peaceful protest and violent conflict events observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. Conflict and protest data from GDELT (2018). Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B28: Conflict events vs protest events: Alternative sources (ICEWS (2018) data)

Dep. Variable:	(1) *Conflict Episode$_{it}$*	(2) *Protest Episode$_{it}$*	(3) Δ *Protest - Conflict$_{it}$*
*(# Schools / # Children)$_i$ * Years since 1978$_t$*	-0.00161*	0.000827	0.00244**
	(0.000960)	(0.000708)	(0.00106)
Observations	5,780	5,780	5,780
R-squared	0.501	0.437	0.249
District FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Province x Year FEs	Yes	Yes	Yes
Sample Mean	.4	.16	-.24

NOTE: The unit of observation is a district i and year t. The sample covers 289 districts across 26 provinces over the period 1995-2014. LPM estimates are reported in all columns. In column 1 the dependent variable is a dummy that takes a value of 1 if a violent event was observed in district i and year t. Column 2 has as dependent variable a dummy that takes a value of 1 if a peaceful protest was observed in district i and year t. In column 3 the dependent variable is the difference between peaceful and violent events observed in district i and year t. The variable *(# Schools/# Children)$_i$* represents the number of primary schools constructed under the INPRES program per 1,000 school-aged children in a district i. The variable *Years since 1978$_t$* is a measure that until 1978 takes value 0, in 1979 takes value 1, in 1980 takes value 2, and so on. Conflict and protest data from ICEWS (2018). Robust standard error clustered at the district level are reported in parenthesis. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.