Schools and Violence:

Regression Discontinuity Evidence from Pakistan

Sultan Mehmood* Georgii Marinichev Faiz Ur Rehman

New Economic School MIT IBA Karachi

August 2026

Abstract

School construction is commonly evaluated by its effects on educational attainment and later-life economic outcomes. We show that its educational gains can coexist with violent political consequences. We study Pakistan's expansion of primary schooling in the 1980s. Beginning in 1983, a government rule made villages with at least 300 residents eligible for a public school. We digitize village-level census records and exploit this sharp threshold in a regression discontinuity design, comparing otherwise similar villages just above and below the cutoff. Eligible villages experience more terrorist attacks once cohorts exposed to the school-construction program reach ages at which political mobilization becomes possible, with no corresponding increase before construction or in the years immediately following it. The program nonetheless achieves its educational objective. In individual-level data, exposed cohorts are more likely to attend school and acquire basic reading, writing, and numeracy skills. At the village level, we also find no detectable increase in local madrassa expansion at the eligibility cutoff or in the likelihood of attending *deeni madaris* (religious schools). These results suggest that a successful expansion of public schooling can build human capital while also generating violent political externalities.

JEL Codes: D74, I25, H52

*Authors are listed in reverse alphabetical order by given name and contributed equally. We thank Jesse Shapiro, Irina Denisova as well as seminar and conference participants at NES and ETH Zurich for their helpful comments and discussions.

1. Introduction

Can building a school make a community more educated and more violent at the same time? Public-school construction is among the most familiar investments in development: by expanding access to education, it can raise schooling, build skills, and improve later-life opportunities (Lagakos et al., 2018).[1] We show that these intended human-capital gains can coexist with a very different political consequence: increased political violence. In Pakistan, villages made eligible for public-school construction become more educated, but also experience more terrorist attacks once the children exposed to the reform reach ages at which political mobilization becomes plausible.

This possibility is not implied by existing evidence on either education or conflict. A literature on militant selection shows that participants in terrorist organizations are not disproportionately poor or uneducated. If anything, participation is often positively associated with schooling and economic status. Krueger (2007) documents this pattern across settings, while Berrebi (2007) finds similar evidence using biographical data on Hamas and Palestinian Islamic Jihad operatives. Bueno de Mesquita (2005) provides a natural explanation: militant organizations may select higher-ability recruits, so more educated individuals can be overrepresented among participants even when deprivation expands the broader pool of potential recruits. A separate literature shows that adverse economic shocks can increase political violence (Miguel et al., 2004).[2] Yet neither literature identifies what happens when the state itself expands access to schooling.

Answering this question is difficult for three reasons. First, schools are rarely built at random: educational expansion typically occurs in places that differ in development, state presence, political importance, or prior exposure to conflict. Second, the relevant outcomes may be extreme and delayed. Existing work shows that schools can act as a tool of political persuasion and can shape attitudes, identities, and voting behavior (DellaVigna and Kaplan, 2007; Fouka, 2020; Adena et al., 2015; Cantoni et al., 2017; Bandiera et al., 2019; Bazzi et al., 2025). Much less is known about whether educational expansion can generate delayed effects on violent political activity, such as terrorism, especially once exposed cohorts reach mobilization age. If schools alter mobilization through the cohorts they educate, violence should not rise when construction begins, but when exposed children reach adolescence and adulthood. Third, an increase in violence must be distinguished from policy failure. If a school-building program does not increase schooling or skills, any subsequent conflict could reflect failed implementation rather than a consequence of successful educational expansion. Establishing the relevant tension therefore requires exogenous variation in school construction, a sufficiently long outcome horizon,

[1] A canonical example is Duflo (2001), who studies Indonesia's large public-school construction program and documents gains in educational attainment and wages among exposed cohorts. Our analysis examines an additional consequence of school construction: whether it can generate violent political externalities as exposed cohorts reach adulthood.

[2] See also Angrist and Kugler (2008) and Dube and Vargas (2013) on commodity prices and violence in Colombia; Bazzi and Blattman (2014) on export price shocks and conflict across countries; and Jia (2014) on droughts and peasant revolts in historical China.

and direct evidence that the program increased education among the cohorts it reached.

We provide such evidence from Pakistan's expansion of public primary schooling under the Sixth Five-Year Plan. Beginning in 1983, Pakistan used population recorded in the predetermined 1981 Population Census to target rural villages for public-school construction: villages with at least 300 residents became eligible for a public primary school. The rule was implemented in a context of large unmet demand for basic education. A UNESCO background study describes Pakistan's education sector as characterized by low literacy, low enrollment, and large rural–urban disparities, and notes that successive education policies had long emphasized universal primary schooling while implementation shortfalls repeatedly pushed target dates forward (Mujahid-Mukhtar, 1999). The reform therefore represented a substantial expansion of public primary schooling into underserved rural communities.

We digitize and geolocate village-level census records across Pakistan and exploit the 300-resident eligibility threshold in a regression discontinuity design. Villages just above and just below the cutoff are similar in pre-reform population and predetermined characteristics, but differ in eligibility for public-school construction; density and balance tests provide no evidence of manipulation around the cutoff. Because the program itself used this population rule to allocate new schools, the comparison around the cutoff captures the consequences of the rule through which rural school construction was expanded. We link these villages to subsequent terrorist attacks and madrassa construction and complement these data by recovering and digitizing 132 Urdu- and English-language primary-school textbooks—approximately 10,000 pages across Grades 1–5—from the reform period.

The schooling expansion was accompanied by meaningful gains in education. Using nationwide household survey data covering more than 470,000 person-level observations, we show that cohorts exposed to the rural schooling expansion are more likely to attend school and acquire basic skills. Exposure raises the probability of reading a simple sentence by approximately five percentage points, writing by roughly one percentage point, and solving a basic numerical problem by approximately three percentage points. The schooling expansion therefore generated meaningful improvements in human capital among the cohorts it reached.

These educational gains were accompanied by a substantial increase in subsequent political violence. Villages just above the 300-person cutoff experience significantly more terrorist attacks during 1993–2002 than villages just below it. Because school construction began in 1983, this is the first decade in which children exposed to the reform begin reaching adolescence and adulthood, when participation in political mobilization and armed organizations becomes plausible. The increase in violence persists, suggesting that the local conflict consequences of exposure to public schooling were not confined to the initial coming-of-age period. These estimates should not be interpreted as showing that treated students became terrorists, since we do not observe individual recruitment. Rather, the design identifies a local increase in political violence following quasi-random exposure to public schooling, which emerges when exposed cohorts come of age and continues over the medium term. Consistent with the design,

terrorist attacks exhibit no discontinuity at the eligibility threshold before the schooling expansion began, indicating that villages on either side of the cutoff were not already on different conflict trajectories prior to treatment.

The central finding is therefore not that schooling failed, nor that poorly educated communities became violent. A public-school construction program that successfully increased schooling and basic skills also generated a politically consequential externality that conventional evaluations of educational expansion would miss.

The composition of the violence is also informative. The largest positive estimates are for attacks attributed to domestic jihadist organizations, while those for separatist, party-political, tribal, and Afghan-war spillover violence are generally smaller and less precisely estimated. The increase in terrorism appears concentrated in domestic jihadist violence rather than spread uniformly across perpetrator types. Because these organizations mobilize around religious and political claims, this pattern makes the content of the schooling expansion particularly relevant to examine.

We find no detectable increase in either individual madrassa attendance or local madrassa expansion, helping distinguish the result from a madrassa-based channel. This shifts attention toward what children encountered within the public schools themselves. Contemporary critiques of Pakistan's public curriculum similarly cautioned that "madrassas are not the only institutions shaping exclusionary worldviews" and identified incitement to militancy and violence in official curricula and textbooks (Nayyar and Salim, eds, 2003).[3]

To characterize this exposure more directly, we recover and digitize 132 primary-school textbooks from the reform period, comprising approximately 10,000 pages. The resulting curriculum data show substantial variation in references to violence across subjects and grades. Such references are most prevalent in Islamiat, but are also present in Urdu and social-studies textbooks, while appearing much less frequently in mathematics and science. Violent content also becomes more prevalent in higher grades. Thus, the schooling expansion increased exposure not only to basic-skills instruction, but also to religious, civic, and historical material in which conflict and violence were sometimes salient.

This evidence suggests one possible link between the schooling and violence results. Because the reform increased public-school attendance, it also increased exposure to the curriculum delivered through those schools. The concentration of conflict-related material in subjects concerned with religion, language, and society, alongside the subsequent rise in domestic jihadist violence, is suggestive of a role for schooling in shaping political and religious narratives relevant to later mobilization. The textbook analysis thus helps characterize the broader social and political content of the schooling treatment experienced by exposed cohorts.

More broadly, public schools may shape political behavior not only through the skills and

[3]Extensive fieldwork on Pakistan's madrassa sector emphasizes substantial institutional heterogeneity and rejects simple accounts based on poverty or automatic radicalization, instead highlighting the social and institutional reasons why families, students, and religious elites engage with madrassas (Bano, 2012).

content they provide, but also through their place in local social organization. Schools extend the state into local social life and give it a role in the formation of future citizens. In settings where education and moral authority are also organized through non-state institutions, expanding public schooling may therefore alter competition over local authority even when households value the educational opportunities it provides. The same intervention can thus build human capital while changing the political environment in which that human capital is produced.

Related Literature. First, we contribute to research on the economic and institutional determinants of political violence. A large body of literature shows that economic conditions shape conflict. Miguel et al. (2004) use rainfall-driven income shocks to show that negative growth increases civil conflict in sub-Saharan Africa. Subsequent work documents related effects of commodity-price and resource shocks in Colombia (Dube and Vargas, 2013; Angrist and Kugler, 2008), export-price shocks across countries (Bazzi and Blattman, 2014), and fiscal austerity (Fetzer, 2019).[4] If schooling affected conflict only by raising skills and future incomes, this literature would suggest that public-school construction should reduce violence by increasing the opportunity cost of joining armed organizations. Instead, we find that violence rises after the reform successfully expands schooling and learning. Our results therefore identify a political consequence of educational expansion that is not reducible to failed human-capital formation.

Our setting is also related to work showing that development interventions can become politically contested. Closest in design, Crost et al. (2014) use a regression discontinuity around a poverty-eligibility threshold in the Philippines and find that barely eligible municipalities experienced greater insurgent violence, consistent with armed groups attempting to undermine a program that could strengthen the state. Nunn and Qian (2014) similarly find that US food aid increased the incidence and duration of civil conflict. We differ in studying a human-capital intervention and in tracing its consequences over a horizon long enough for exposed children to reach mobilization age. The timing matters: we find no increase in attacks before the program or in the period immediately after school construction, but a rise once exposed cohorts reach adolescence and adulthood that persists into the medium term.

Second, we speak to research on education, political socialization, and militant mobilization. At the individual level, Krueger and Malečková (2003) show that support for and participation in militant activity do not decline with education or living standards, while Berrebi (2007) documents positive selection on schooling among Palestinian militants. Bueno de Mesquita (2005) provides a natural explanation: militant organizations may select higher-ability recruits, making educated individuals overrepresented among participants even if deprivation increases the pool of willing volunteers. At the institutional level, schooling can shape political identities as well as skills. Cantoni et al. (2017) show that curricula causally affect students' political

[4]The local-level evidence is more nuanced than a single income channel. McGuirk and Burke (2020) and Harari and La Ferrara (2018) find that agricultural price and weather shocks raise some forms of violence while reducing others, depending on whether the relevant income is labor-based or appropriable, while Berman et al. (2017) show that higher mineral prices intensify local conflict over resource rents.

attitudes, while Alesina et al. (2021) and Bandiera et al. (2019) emphasize the role of primary education in forging national attachment. Conversely, Fouka (2020) shows that school-based assimilation policies can harden minority identity rather than dissolve it. In settings where public and religious schools coexist, as in Pakistan (Andrabi et al., 2006), the expansion of state schooling may therefore reshape political mobilization even while improving educational attainment.

Third, we contribute to work on the long-run consequences of childhood exposure to public investment. Cunha and Heckman (2007) and Cunha et al. (2010) emphasize self-productivity and dynamic complementarity in skill formation, while Heckman et al. (2006) shows that childhood cognitive and noncognitive skills predict adult outcomes beyond the labor market. We extend this logic to political violence by showing how early exposure to primary schooling can shape later political behavior. Terrorist attacks do not rise when schools are built, but increase once exposed cohorts reach mobilization age.

Finally, our paper connects the literature on public-school expansion to evidence on educational institutions and religious political mobilization. A central finding in development economics is that large-scale school-construction programs increase educational attainment and improve later-life outcomes. In particular, Duflo (2001) exploits Indonesia's INPRES program to show that exposure to newly built primary schools increased schooling and wages among exposed cohorts.[5] A separate literature shows that religious educational and political institutions can generate durable forms of Islamist mobilization: Bazzi et al. (2020) study Islamic institutions in Indonesia, Meyersson (2014) studies Islamic local rule in Turkey, and Benzer and Tukiainen (2025) shows that the expansion of state-run religious secondary schools in 1970s Turkiye strengthened Islamist youth organizations and later increased Islamist vote share.

Our setting differs from both literatures. The contribution is not that school construction failed to build human capital in Pakistan, nor that religious schooling mobilized Islamist politics. Rather, we show that even a successful expansion of public schooling can generate politically consequential externalities that conventional evaluations of educational expansion would miss.

2. Institutional Background

Colonial Origins. Pakistan was created in 1947 through the partition of British India, alongside independent India, and began its independent history with limited access to public primary education, especially across its predominantly rural population. This inherited educational deficit was consequential for a new state seeking to govern geographically dispersed and socially diverse

[5] Subsequent work using the INPRES design has documented effects on wages and formal employment (Duflo, 2004), adult health and children's human capital (Mazumder et al., 2023), long-run employment, migration, marriage-market outcomes, living standards, tax payments, and intergenerational educational transmission (Akresh et al., 2023), as well as heterogeneous female-schooling responses across marriage-payment institutions (Ashraf et al., 2020). This literature has focused primarily on educational, labor-market, health, fiscal, and family outcomes. Also related is, Bazzi et al. (2025), who show that INPRES public school construction increased overall Islamic secondary school enrollment and left exposed cohorts more invested in religious identity.

regions. Expanding primary schooling therefore became central to Pakistan's development project: schools promised to raise literacy and basic skills while extending a visible and durable state institution into village life. A public school was to be a claim by the state over the education and socialization of children in communities where these functions could already be shaped by religious authorities and other non-state institutions (Bengali, 1999).

Pakistan's Educational Deficit and Schooling Landscape. The educational deficit remained severe decades after independence. In the early 2000s, Pakistan's adult literacy rate was 44 percent, compared with a South Asian average of 54 percent, and its net enrollment rate was 51 percent, compared with 83 percent in India, 90 percent in Sri Lanka, and 70 percent in Nepal (Andrabi et al., 2008). Yet Pakistan's educational landscape cannot be reduced to a simple public-school versus madrassa divide. Although religious schooling became prominent in policy discussions about radicalization, madrassas accounted for less than one percent of total enrollment, with no evidence of a dramatic contemporaneous increase. The major expansion outside the public system was instead in low-fee private schooling: in 2000s, private schools enrolled 35 percent of primary-school students, including substantial numbers of rural and lower-income children (Andrabi et al., 2008). This distinction matters for our interpretation. We do not assume that public-school entry mechanically displaced children into religious schools; rather, we ask whether state expansion in public schooling itself may have radicalized the population to political violence.

Development Planning and Primary Schooling. Pakistan pursued this objective through a sequence of national development plans. The First Five-Year Plan treated universal primary education as essential to citizens' economic and civic participation, while subsequent plans continued to make school access a stated policy priority (National Planning Board, Government of Pakistan, 1957). Yet the expansion of public education remained incomplete. Rural access was limited, and the public system continued to face serious deficiencies in coverage, effectiveness, and quality (Bari et al., 2005). The Sixth Five-Year Plan, implemented beginning in 1983, represented a renewed effort to expand primary education in rural Pakistan (Planning Commission, Government of Pakistan, 1983). It is this expansion that provides the setting for our analysis.

Public Schooling and Social Content. Public schooling can build skills while also transmitting political and social content. This channel is plausible in Pakistan because the curriculum was centralized. The SDPI report describes a system in which the Curriculum Wing formulated national curricula and guidelines, while provincial textbook boards authorized books for public schools (Nayyar and Salim, eds, 2003). A second curriculum report similarly notes that the Curriculum Wing prepared the "National Curriculum Document" and "Guidelines," and that provincial textbook boards then produced and distributed textbooks (Center for Research and Security Studies, 2008). These reports caution against locating ideological socialization only

in madrassas: SDPI identified "incitement of militancy and violence" and "the glorification of wars and the use of force" in official curricula and textbooks, and notes that government schools enrolled nearly 20 million children, compared with 1.5 million in madrassas.

Political Violence and Terrorism in Pakistan. Pakistan's modern experience with political violence developed in distinct waves. The late 1970s and 1980s were formative. General Zia-ul-Haq's Islamization coincided with the anti-Soviet jihad in neighboring Afghanistan, a period in which Islamist organizations, armed networks, and sectarian mobilization expanded substantially. These forces persisted into the 1990s, alongside militant organizations oriented toward sectarian conflict and Kashmir (Nasr, 2000). This history is particularly relevant for our setting because the public-school expansion we study began in 1983, during the same period in which the state was expanding schooling while also reshaping the political and religious content of public institutions. The resulting historical overlap has long linked education and militancy in discussions of Pakistan (Haqqani, 2005). After 9/11, violence entered a different phase. Pakistan's involvement in the U.S.-led war in Afghanistan, subsequent military operations in the northwest, and the emergence of the Pakistani Taliban shifted an important part of militant violence toward the Pakistani state and domestic targets. Suicide bombing, relatively uncommon in earlier periods, became a prominent tactic during the 2000s, with a particularly sharp escalation later in the decade (Pape and Feldman, 2010). The violence observed in our data therefore does not represent a single conflict or organization, but successive episodes involving different militant groups, objectives, and tactics. Our empirical measure follows the widely used Global Terrorism Database (GTD) definition of political violence, which records intentional acts or threats of violence by non-state actors pursuing political, economic, religious, or social objectives through coercion or intimidation (LaFree and Dugan, 2007).

Schooling Expansion and the Assignment Discontinuity. In the school-expansion program we study, eligibility followed a sharp assignment rule based on the predetermined 1981 Population Census: villages with at least 300 residents were targeted for public primary school construction. We digitize village-level census records, geolocate settlements across Pakistan, and exploit this threshold in a regression discontinuity design. Villages just above and below the cutoff were similar in initial population and predetermined characteristics, but differed in their eligibility for state-school construction. While subsequent reports indicate that a substantial share of the mandated schools were in fact constructed Planning Commission, Government of Pakistan (1988), village-level data identifying which specific villages received a school under the construction rule are, to the best of our knowledge, unavailable. Our baseline estimates therefore identify the reduced-form effect of eligibility for the schooling expansion on subsequent violence. This estimate remains policy-relevant: a policymaker designing such a program would be interested precisely in this intention-to-treat effect, since ex ante they cannot anticipate the delays and implementation failures that determine which eligible villages ultimately receive a school.

3. Data and Descriptive Statistics

This study draws on several sources: modern and historical censuses and individual-level surveys. The main results use three datasets: a village-level dataset of terrorist attacks, a merged village-level dataset from the 1981 and 2020 Censuses, and an individual-level survey of educational outcomes.

Education and Terrorist Attacks Measuring the effect of education on terrorism requires a dataset that links village-level population to terrorist attacks occurring near each village. We construct one by digitizing the 1981 Census, which lists the rural localities in each district with their population counts. The 1983 school construction program was based on the 1981 Census, the most recent available at the time. We then geolocate the villages by district and village name, assigning each a set of coordinates. Figure A1 reports the map of villages, we were able to assign the coordinates to. These let us spatially merge the village-level dataset with the Global Terrorism Database National Consortium for the Study of Terrorism and Responses to Terrorism (2020), which covers terrorist attacks worldwide from 1970 to 2021 as well as their location, casualties and timing. Panel A of Table A1 reports summary statistics for this dataset and the main variables used in the analysis. Appendix B. Dataset Construction details each step of the dataset construction. In total, we ended up with 13000 villages across Pakistan with available coordinates with 2,231 village below the threshold of 300 people in comparison group.

Matched Census Dataset To examine the long-term consequences of the reform and the response by government and society, we link the 1981 Census rural localities to the 2020 Mouza (Village) Census of Pakistan (Pakistan Bureau of Statistics, 2020), matching on village name and on 1981 districts cross-walked to modern ones to account for boundary changes. This procedure yielded a dataset of more than 33,000 villages matched to the 2020 census, with nearly 7,000 villages having populations below 300 and forming the comparison group. The 2020 Census records general village characteristics along with educational and public amenities, letting us measure effects on public goods and civil society. Panel B of Table A1 reports summary statistics for this dataset; Appendix B. Dataset Construction discusses the construction process in more detail.

Individual-Level Survey. To assess whether the reform achieved its goal of raising educational enrollment across rural Pakistan, we use the Pakistan Social and Living Standards Measurement (PSLM) survey (Pakistan Bureau of Statistics, 2021). It tracks educational outcomes, including school and madrassa attendance and the ability to read, write, and count, by birth cohort from 1952 to 2001. Panel C of Table A1 reports summary statistics for the individual-level variables in this dataset.

4. First Step: Effect on Educational Outcomes

We begin by documenting that the 1983 school construction program increased educational attainment (Planning Commission, Government of Pakistan, 1983). This exercise serves as a first-stage validation of the reform: if the program expanded access to schooling, cohorts exposed to the reform should exhibit higher school enrollment and stronger basic skills, especially in rural areas where the school-building program was targeted.

We use individual-level data from the Pakistan Social and Living Standards Measurement Survey (PSLM), which reports school attendance and basic skills by birth cohort. The empirical design compares rural and urban respondents within the same district across cohorts born before and after the reform. The estimating equation is:

$$y_{it} = \beta \times \text{Rural}_i \times \text{Post}_t + \gamma_{dr} + \gamma_t + \gamma_s + \Gamma' X + \varepsilon_{it} \tag{1}$$

Here, i indexes individuals, t indexes birth cohorts, d indexes districts, r indexes rural or urban residence, and s indexes survey years. The outcome y_{it} is an educational outcome, such as school attendance or the ability to read, write, or count. The indicator Rural_i equals one for respondents living in rural areas. The indicator Post_t equals one for cohorts born in 1978 or later, who were young enough to enter primary school after the reform began.[6] The vector X_i includes individual controls for gender and age. The specification includes district-by-residence fixed effects, γ_{dr}, birth-cohort fixed effects, γ_t, and survey-year fixed effects, γ_s. Standard errors are clustered at the district level.

The coefficient β captures the differential change in educational outcomes for treated rural cohorts relative to the corresponding within-district comparison group. This comparison nets out permanent differences across rural and urban areas within districts, as well as common cohort and survey-year shocks.

The identifying assumption is that, absent the reform, rural and urban cohorts within the same district would have followed parallel trends in educational outcomes. To assess this assumption, we also estimate a dynamic version of equation (1) and examine whether rural and urban areas followed similar pre-reform trends.

Figure 1 reports the dynamic first-stage results. The pre-reform coefficients are close to zero, supporting the parallel-trends assumption. After the reform, school attendance rises for treated rural cohorts relative to urban cohorts. The estimates become larger for later cohorts, consistent with the time required to build, staff, and equip schools, as well as with documented implementation delays (Saleem, 2022). By the 2001 cohort, program eligibility increased school attendance in rural areas by approximately 10 percentage points. This result provides a first-stage validation: the expansion of village schools increased attendance among the cohorts exposed to the program. We next examine whether this increased access also produced broader

[6] A child born in 1978 would be 6 in 1983, the year the reform began.

human-capital gains, as measured by the ability to read, write, and count. These outcomes examine whether increased schooling exposure translated into basic human-capital accumulation, which is a central objective of the primary schooling expansion.

Table 1 reports the results. The coefficients are positive and significant in every case. Reading ability rises by 9% relative to the mean, while writing and counting rise by 1 and 4 percent respectively.

These results indicate both the scale of the reform and its success: the expansion raised school enrollment and basic skill acquisition, consistent with the literature (Duflo, 2001).

They also matter because data limitations prevent us from estimating the first-stage impact on school construction directly. To our knowledge, data on the number of schools built and their precise locations do not exist, and some of the schools are no longer operating (HBond, 2025). We therefore report reduced-form effects throughout the following analysis, and the estimates should be interpreted accordingly.

5. Second Step: Effect on Political Violence

We next turn to the village-level regression discontinuity design. This analysis asks whether exposure to the school-construction reform affected subsequent terrorist attacks. We estimate the following specification at the village level:

$$y_i = \alpha + \beta \times \mathbb{1}[\text{Population}_i \geq 300] + f(\text{Population}_i) + \mathbb{1}[\text{Population}_i \geq 300] \times g(\text{Population}_i) + \varepsilon_i \tag{2}$$

where i indexes a village and β, the coefficient of interest, measures the discontinuity at the population threshold of 300. The functions f and g control flexibly for the running variable on either side of the threshold; we approximate them with local polynomials. The baseline uses local linear regression, and we report robustness to higher-degree polynomials. y_i is an outcome of interest, α is a constant and ε_i the error term.

We estimate this regression using the robust estimator of Calonico et al. (2014). Specifically, we estimate the regression on selected MSE-optimal bandwidth and reweight observations using triangular kernel weights based on the proximity to the threshold. Since conventional inference is suboptimal under the proposed estimator, as a baseline, we report heteroskedasticity-robust standard errors based on a nearest-neighbor variance estimator, developed to perform well in our setting (Calonico et al., 2018).

Imbens and Lemieux (2008) identify manipulation of the running variable and confounding discontinuities as the two main threats to the validity of a regression discontinuity design. Eggers et al. (2018) restate these for population-based designs: sorting, where localities manipulate their recorded population, and compound treatment, where other policies change at the same threshold. In our case however, both concerns are unlikely to be present.

First, school placement was based on the 1981 Census, two years before the Sixth Five Year Plan introduced the policy. Because the assignment variable was recorded before the policy existed, villages could not have anticipated the 300-resident threshold or adjusted their reported population to fall above it. This timing rules out sorting by construction. We nonetheless test for it directly with two density tests of the running variable. Under sorting, the density of villages would jump at the threshold: villages would be discontinuously more likely to report 300 residents than 299, as those just below the cutoff misreport to qualify for a school. The McCrary (2008) and Cattaneo et al. (2020) tests, shown in Figure 2, do not reject continuity of the population density at the threshold, consistent with the absence of manipulation.

Second, although we cannot rule out a confounding policy entirely, one is unlikely. Nationwide programs of comparable scale are documented in the Five Year Plans, and none shares the 300-resident threshold. Other population-based rules exist, such as the placement of Rural Health Centers, but their cutoffs lie far from ours, e.g., 10,000 residents for Rural Health Centers, well outside the MSE-optimal bandwidth in our analysis.

Figure 3 reports tests on the village characteristics available in the 1981 Census. We find no discontinuity at the threshold in baseline literacy, geographic location, or religious composition. The absence of pre-treatment discontinuities is consistent with valid identification: villages just above and below the cutoff are comparable on observables.

We now estimate whether eligibility for public-school construction changed subsequent political violence. The outcome is the log number of terrorist attacks within 20 kilometers of a village in each ten-year period. The specification is the regression-discontinuity design described above, with MSE-optimal bandwidths and local polynomial controls on either side of the 300-person cutoff.

Before estimating the main effect, Figure 4 reports the regression on the outcome in the pretreatment period. Panels A and B show no discontinuity in attacks before the reform, 1973–1982, or in the first decade after the reform began, 1983–1992, when the schools were constructed. The absence of an immediate jump is informative because schools were assigned beginning in 1983. If the result reflected pre-existing violence, direct targeting of newly eligible villages, or a mechanical disruption caused by construction, one would expect a discontinuity either before the reform or soon after it began.

The discontinuity appears in Figure 5. Panel A reports the result for 1993–2002. This is the first decade in which children exposed to the school-construction program begin reaching adolescence and early adulthood. Villages just above the 300-person eligibility cutoff experience more terrorist attacks than villages just below it. Panel B shows that the increase persists in 2003–2012. The timing is consistent with an effect operating through exposed cohorts rather than through the construction event itself.

Table 2 reports the corresponding estimates. The coefficient is small and statistically insignificant before the reform. It remains statistically insignificant in 1983–1992. It becomes positive and significant in 1993–2002 and remains positive and significant in 2003–2012. The

estimates imply that eligibility for public-school construction raised later terrorist activity near treated villages once exposed cohorts reached mobilization age.

The magnitude of the effect is economically significant. Eligibility for school construction increased the number of terrorist attacks by 39% in the subsequent decade and by 52% in the following one. As with any RD design, this estimate identifies a local average treatment effect for complier villages near the threshold. Even under this narrower interpretation, however, the results underscore the importance of education and value transmission as a central channel of radicalization and political violence in Pakistan.

This evidence should be interpreted as a village-level effect of eligibility for school construction, not as evidence on individual recruitment. We do not observe whether a particular student later joined an armed group. The design instead identifies whether villages assigned to receive the public-school intervention later experienced more political violence. The result is striking because the same eligibility rule increased schooling and basic skills. The increase in terrorism is not a symptom of failed educational expansion; it appears after the reform expanded human capital.

Table 3 shows that the increase is not limited to event counts. The estimated discontinuity also implies additional fatal casualties in the post-exposure periods. Aggregating the village-level estimates over eligible villages, we construct a counterfactual by assuming the same percentage effect for every treated village. Our back of the envelope calculations show that the reform implies roughly 48000 additional deaths in 1993–2002 and about 200000 additional deaths in 2003–2012, roughly 40% of all deaths in our sample. These magnitudes suggest that the post-exposure discontinuity corresponds to substantively meaningful violence.

Figure 6 decomposes the main result by perpetrator. We assign every group named in GTD [7] to one of seven clusters and re-estimate the main specification with the log number of attacks by cluster as the dependent variable.

The effect is concentrated among domestic jihadist groups, where it is largest and statistically significant. This is what the recruitment channel predicts: these organizations recruit locally and are best placed to absorb recent school leavers.

No other cluster shows a consistent effect. Coefficients are insignificant in both periods for international jihadist organizations such as al-Qaida and the Islamic State, for Baloch and other separatist groups, and for party-political and tribal or criminal violence.

Attacks by groups party to the war in Afghanistan follow a different pattern: the coefficient is insignificant in 1993–2002 and significant in 2003–2012. The timing tracks the post-2001 intensification of the Afghan insurgency. Recruitment networks linking Pakistani seminaries to that conflict were built during the anti-Soviet war and predate the Taliban, which inherited rather than created them (Abou Zahab and Roy, 2004).

[7] While a significant portion of attacks does not have a recorded perpetrator, we believe that a subsample analysis here still provides an important perspective on the nature of the observed effect.

Robustness. The main result is stable across the usual RD diagnostics and implementation choices. We first address the two standard concerns in population-based RD designs. As discussed above, the running variable was recorded in the 1981 Census before the school-construction rule was introduced, and the density tests in Figure 2 show no evidence of sorting around the 300-person cutoff. Predetermined village characteristics are also smooth at the threshold, as shown in Figure 3.

We then assess whether the post-exposure result depends on bandwidth choice or functional form. Figure A3 reports bandwidth sensitivity for terrorist attacks in 1993–2002 and 2003–2012. The estimates remain positive, stable and significant over the relevant range of bandwidths spanning from half to twice of MSE-optimal bandwidth. Figure A4 repeats this robustness check for casualties result. All reported estimates are similar to baseline and significant.

Figure A5 replaces baseline local linear regression with the specification with quadratic polynomial on the both sides of the threshold. The estimates yielded by this approach are similar in magnitude to the baseline and significant. Figure A6 does the same for casualties and these results also remain robust.

Table A2 reports the corresponding regression estimates under alternative outcome transformations. Following recent literature on imputation of zeros under log transformation (Chen and Roth, 2024) we show that our results remain positive and significant under different ways to deal with zeros in log specifications as well as other functional form of the outcome.

In the process of construction of our dataset we faced several choices, such as choosing the geographical buffer to match attacks to villages or for which time periods to report the results. All these choices might have affected the validity of our results. Appendix B. Dataset Construction reports several robustness exercises regarding the creation of our variables. Our results remain robust to variety of such choices, showing that the main findings do not depend on our arbitrary choice.

The timing also supports the design. The discontinuity is absent before the reform and in the decade immediately after construction begins, but appears when exposed cohorts reach adolescence and early adulthood. Table 3 shows the same pattern for casualties: the post-exposure periods display positive effects on fatal and wounded casualties, while the pre-reform period does not.

These checks support the interpretation of the main result. The increase in violence is not driven by manipulation of the cutoff, sharp imbalance in predetermined village characteristics, or a particular bandwidth or polynomial choice. The evidence points to a delayed increase in political violence in villages made eligible for public-school construction.

6. Mechanisms

Public primary schools can change more than access to classrooms. They also brought children into a common state curriculum and a new institutional setting for social and political formation.

This channel is especially relevant in Pakistan, where curriculum content was centrally organized, and public schools were the main mass educational institution.

School Curriculum. The most direct exposure of minds and attitudes of children to public schools is conveyed through what they learn. To shed light on this we examine the content of primary-school textbooks to characterize what the schooling expansion brought to children beyond basic literacy and numeracy. We recovered and digitized 132 Urdu- and English-language primary-school textbooks from the relevant reform period, when children would be exposed to this reform, covering roughly 10,000 pages across Grades 1–5. [8] This is the relevant curriculum because the assignment rule expanded eligibility for public primary schools. The page is the unit of analysis.

For each page, we use Claude Sonnet 5 to classify whether the text contains explicit Islamic content, such as references to Islam, Muslims, the Prophet, the Quran, Islamic practices, Islamic figures, or Islamic history; or violent content, for example, references to wars, physical violence, punishments or martyrdom. Appendix B. Dataset Construction reports examples of the digitized textbook material, and Appendix A3 reports the exact prompts used for classification.

Figure A7 shows that Islamic content is pervasive but unevenly distributed. As expected, it is concentrated in Islamiat, a subject for Islamic Studies, where about 70 percent of pages contain Islamic references. But, at the same time, it also appears in Urdu and Social Studies, the subjects through which civic, historical, and moral narratives are often transmitted. The grade profile is also informative: Islamic content rises across primary grades and is highest by Grade 5. Thus, the public-school curriculum exposed children not only to basic skills, but also to religious and civic material embedded in core subjects.

We then apply the same page-level procedure to violent content. The LLM classifies whether a page contains references to physical force, war, battle, killing, punishment, armed conflict, or martyrdom involving violence. Figure 7 shows a similar pattern. Violent content is highest in Islamiat and also appears in Urdu, Social Studies, and English, while it is nearly absent from Mathematics and low in Science. It also increases with grade, with the largest share in Grade 5.

These patterns provide a plausible link between school construction and later political violence. The program raised attendance and skills, but it also expanded exposure to a curriculum containing religious and violent narratives. The evidence does not imply that textbook content alone explains the terrorism result. Rather, it shows that the treated cohorts were exposed to a form of state schooling that combined human-capital production with political and moral socialization. This fits both the timing of the main result: violence does not rise immediately after construction, but when exposed cohorts reach ages at which recruitment and mobilization become feasible. This finding is also consistent with the decomposition by the terrorist group type. Higher exposure to Islamic and violent context would make it easier for

[8] Our sample contains the universe of primary schools textbooks in collection of one of the biggest libraries in Pakistan.

local Islamic terrorist groups to recruit school graduates which would result in increase in attacks by that groups which is observable in the data.

Alternative Explanations. We consider several other channels that could account for the estimated increase in violence. One possibility is that the expansion of public schooling altered the local education market by shifting students toward religious schools, or madrassas. Public debate has sometimes portrayed madrassas as potential sites of radicalization, at times using the pejorative label "hate factories" (Flynn, 2016), although the broader academic literature provides little support for such a simple link (Andrabi et al., 2006). Figure 8 repeats our cohort-based analysis using madrassa attendance as the outcome and finds no detectable increase among cohorts exposed to the public-school expansion. Figure A9 provides a complementary village-level test, using distance to the nearest religious school in 2020 as the outcome in the population-threshold RDD. We again find no discontinuity at the 300-person eligibility threshold. The evidence provides little indication that the expansion of public schools induced substitution toward religious schooling or a differential expansion of the madrassa network in eligible villages. This contrasts with Bazzi et al. (2025), who find that public primary-school construction in Indonesia increased subsequent enrollment in Islamic secondary schools, as religious schools absorbed part of the additional demand generated by expanded primary education. One reason the same adjustment may be less important in Pakistan is the much smaller size of the religious-school sector: madrassas account for less than 1 percent of national enrollment (Andrabi et al., 2006). Their limited enrollment base leaves considerably less scope for an expansion of public schooling to generate a quantitatively important response through madrassa attendance.

A second possibility is that school construction changed the composition of treated villages by attracting population from elsewhere. Such migration could increase the number of individuals exposed to the risk of violence even in the absence of a change in violence per capita. Direct migration flows are not available at sufficiently fine spatial resolution for the period we study, so we examine whether treated villages exhibit signs of unusually large population growth. Table A3 tests whether treatment increased the probability that a village subsequently became urban or partially urban. We find no statistically detectable effect on either outcome in the baseline specification. Urban status is necessarily a coarse measure of population change, but the absence of differential urbanization provides little evidence of the large and sustained population inflows that would be needed for compositional change alone to account for the main estimates.

Finally, we examine whether the results could reflect differential measurement of violence rather than differential incidence. A village receiving a public school might, for example, become better connected to administrative or media networks, making attacks more likely to appear in the Global Terrorism Database. We use MODIS active-fire detections as a reporting-independent benchmark because satellite measurements of thermal anomalies do not depend on local reporting (Giglio et al., 2016). Since the MODIS record begins in 2000, we restrict this exercise to the overlapping period and match total Fire Radiative Power (FRP) (Kaufman et al., 1998) to villages

over 2003–2012. We then re-estimate the threshold specification using reported attacks relative to local FRP as the outcome. If the estimated discontinuity in terrorism primarily reflected differential reporting around the eligibility threshold, we would expect this measure to shift discontinuously as well. Figure A8 shows no such break at the threshold ($p = 0.695$). While satellite fires are not themselves a measure of terrorist violence, this exercise provides little indication that differential reporting around the eligibility cutoff is driving the estimated effect.

7. Responses by Government and Society.

Educational interventions rarely confine their effects to schooling, and rarely exhaust them within a single cohort's time in school. Both the horizon and the breadth of the response extend well beyond the classroom (Akresh et al., 2023; Clark and Royer, 2013). Our setting is no exception. Linking settlement records from the 1981 Population Census to the 2020 Mouza Census gives us a window of roughly four decades between the pre-program baseline and the outcomes. Over that horizon we find that school construction and exposure to terrorism reshaped conditions in eligible settlements along several margins beyond enrollment and attainment.

Table 4 reports the first of these margins. Treated villages lie on average two kilometers closer to the nearest police station in 2020 than villages on the other side of the eligibility threshold. Since police posts are sited long after the settlements themselves, we read the gap as an adjustment by the state rather than an initial condition. Interpreted this way, enforcement follows violence: the state reallocates policing toward the places where attacks rose. Thus, the reallocation is informative about magnitude. Decisions to establish a post respond to sustained changes in local risk rather than to short-run fluctuations, so a shift of this size indicates that the increase in attacks was large enough to register in administrative allocation.

The response does not extend to health care. Columns 3 and 4 of Table 4 show no detectable change in distance to the healthcare center, with point estimates close to 0 and insignificant. The asymmetry is what one would expect if the state directs each input to where its marginal return is highest. Attacks make up a large share of the incidents that policing addresses, so an increase in attacks shifts the enforcement margin substantially and a post sited nearby raises response capacity where it binds. The same attacks make up a small share of the caseload that health facilities address, the bulk of which is unrelated to violence, so the additional casualties raise demand for care by comparatively little and can plausibly be absorbed by facilities already in place. The two results are consistent only when each is judged against its own baseline, and we treat them as responses on separate margins rather than as two readings of a single response.

The second dimension on which the expansion left a mark is women's participation in organized public life. Table 5 reports the effect on the presence of a women's NGO in the village. This is a particularly informative outcome in Pakistan, where women's mobility and participation outside the household remain closely shaped by local gender norms (Mehmood et al., 2025). Establishing and sustaining a women's organization requires women to meet,

organize, and participate collectively beyond the household, and therefore requires some degree of local acceptance of women's public activity (Dhar et al., 2022). Its presence may also capture a broader willingness to engage with civic organizations. Such trust is consequential in Pakistan: following revelations that the CIA had used a vaccination campaign in the search for Osama bin Laden, militant groups amplified suspicions that ostensibly civilian service providers could serve hidden external agendas, reducing willingness to engage with them in areas with stronger prior Islamist support (Martinez-Bravo and Stegmann, 2022). A women's NGO therefore captures something different from an individual's stated attitude: whether the local environment is sufficiently accepting of both women's collective participation and organized civic activity for such an institution to operate. Treated villages are 1.4 percentage points less likely to host one, a reduction of about eighty percent relative to the mean. We read this as consistent with treated villages remaining more socially conservative and less receptive to organized civic participation four decades after the reform.

Columns 3 and 4 of Table 5 report the same specification with the presence of a citizen board as the outcome. These organizations operate in the same villages and face the same external environment as women's organizations, but their membership and purpose are not gendered. They therefore respond to any force acting on organized civic life in general, whether deteriorating security, weaker state presence, or a decline in collective participation, while remaining unaffected by norms governing women specifically. The estimated effect is close to zero and stable across specifications. We read the contrast in these two estimates as evidence that the effect operates through norms constraining female participation rather than through a general contraction of associational life.

7.1. Robustness.

Figure 9 and Figure 10 present the graphical counterparts to the two results, plotting distance to the nearest police station and the presence of a women's organization against the running variable. Each figure shows together with local linear fits estimated separately on either side of the cutoff and confidence bands. The conditional mean jumps at the threshold in both cases, while away from it the fitted function tracks the binned means closely and displays no comparable break.

We then subject both estimates to the standard sensitivity checks. Figure A11 re-estimates the specification on samples ranging from half to twice the MSE-optimal bandwidth. The point estimates remain stable across this window, and the confidence intervals exclude zero throughout in the most cases, so neither result depends on a particular bandwidth choice.

Figure A10 replaces the local linear fit with second-order polynomials on each side of the threshold. Magnitudes and inference are unchanged.

8. Conclusion.

Terrorism is commonly understood as a phenomenon external to the state, thriving where authority is absent and state capacity is weak (Besley and Persson, 2011). We document the opposite configuration. The Pakistani state did not fail to reach the settlements in our sample. What generated the outcome was action rather than inaction, and the instrument was one ordinarily counted among the state's constructive functions.

The education expansion of the 1980s achieved what it set out to achieve. Enrollment rose, attainment rose, and basic skills improved in rural areas, gains that persist in the data four decades later. Alongside these gains the program carried an ideological curriculum, and along that dimension it was equally effective. Treated settlements subsequently experienced more terrorist attacks, and the associated loss of life runs to tens of thousands. The same institution delivered both outcomes, through the same schools and to the same cohorts.

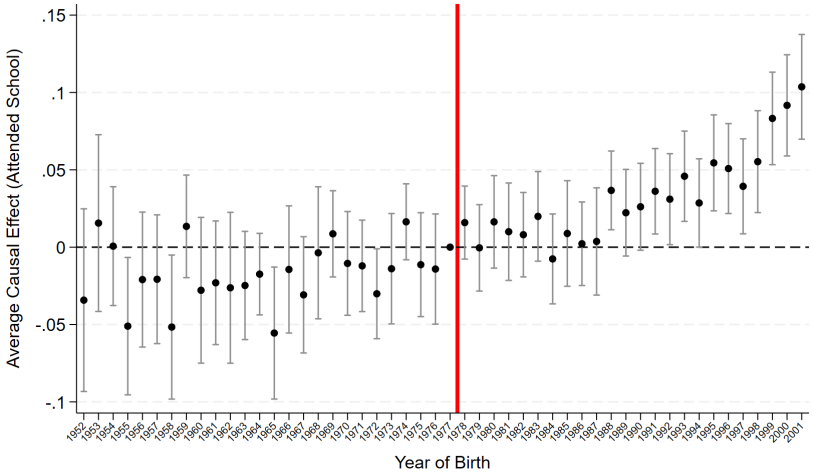
Two implications follow. First, the existing literature has largely rejected deprivation accounts, finding that militants are not drawn disproportionately from the least schooled (Krueger and Malečková, 2003), and has proposed instead that education raises the returns to political participation of all kinds, including violent kinds (Campante and Chor, 2012). Our results identify a channel that operates through content rather than through quantity or returns. This places the finding alongside work showing that curricula shape political attitudes (Cantoni et al., 2017; Voigtländer and Voth, 2015). Our result reaches beyond attitudes to organized violence, and beyond the students themselves to the settlements in which they lived.

The second implication concerns the durability of ideological policy. Zia-ul-Haq's regime ended in 1988, and successive governments revised the curriculum, yet the settlements exposed to the original program remain distinguishable from their neighbors in 2020. What persisted is embedded locally: our evidence indicates that treated settlements sustain fewer organizations supporting women's participation in public life and draw a heavier enforcement presence from the state. The pattern is consistent with a mechanism in which a cohort educated under a particular curriculum transmits its norms to the next generation and to the institutions of the settlement, so that the policy outlives the regime that enacted it. Reversing an ideological program, on this account, requires more than reversing the text.

We close on what these results do and do not recommend. They do not advise against expanding schooling. They are an argument that the content of instruction is a policy instrument of the same order as its quantity, and that it is evaluated far less often. Governments that undertake mass education while a particular ideological project is ascendant bind their successors to consequences that will outlast the project by decades, and the binding is tightest where the state is most effective at delivery.

Figures and Tables.

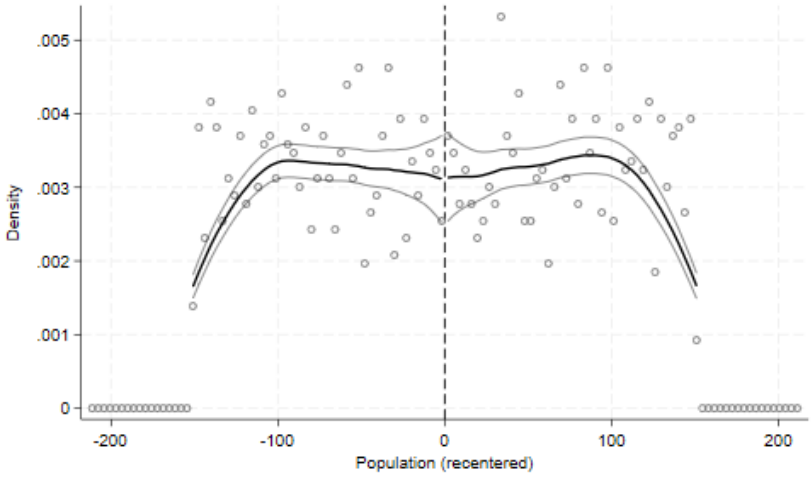
Figure 1: Reform and School Attendance



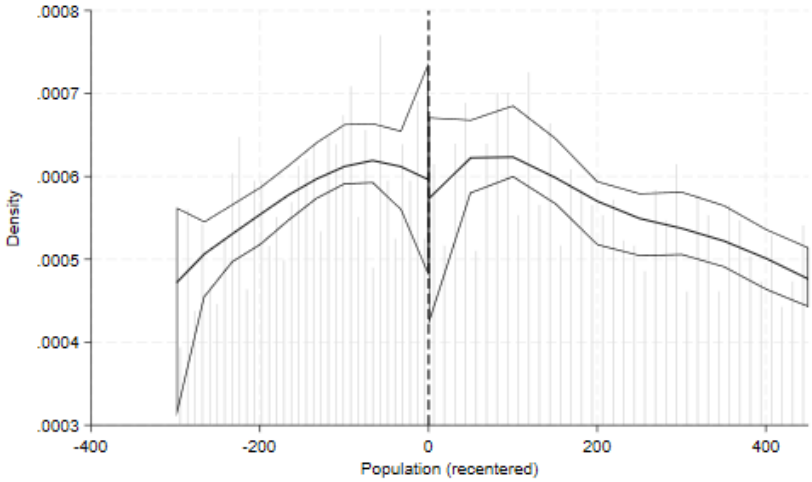
Note: The figure plots cohort-level event-study estimates using individual PSLM respondents. The design compares cohorts born early versus late enough to be school-age after the 1983 school construction reform, separately for rural and urban areas. The dependent variable is an indicator for attending any school. Bars report 95% confidence intervals; standard errors are clustered at the district level. All specifications include district-by-rural/urban, birth-cohort, and survey-year fixed effects.

Figure 2: Continuity at the Cutoff

Panel A: McCrary (2008) Test

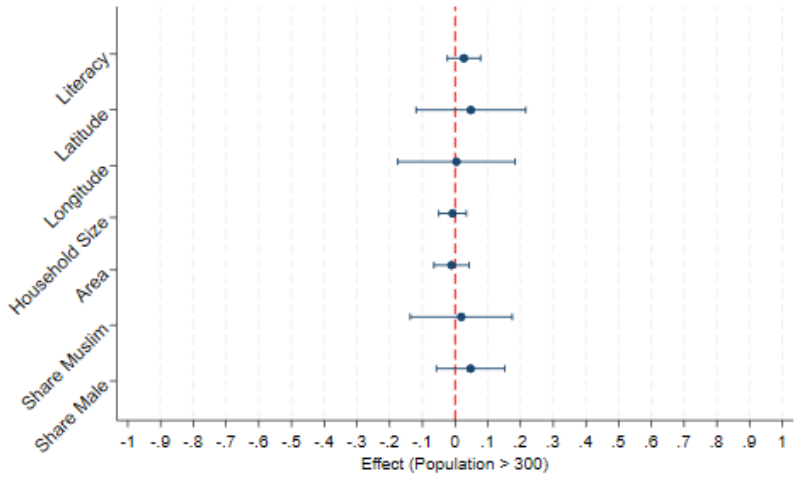


Panel B: Cattaneo et al. (2020) Test



Note: The figure reports the distribution of the village population within the MSE-optimal bandwidth and two density tests. The density is estimated on the both side of the threshold to check the continuity, p-value of McCrary test 0.906 and for Cattaneo et al. is 0.465

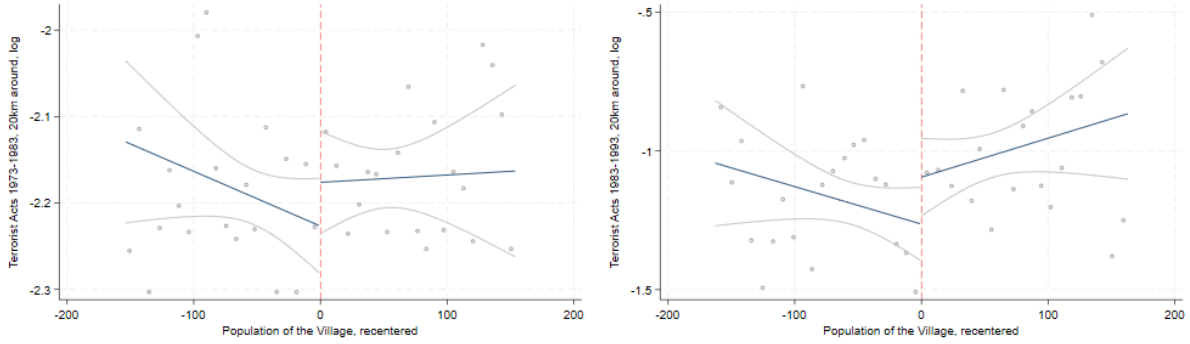
Figure 3: Balance over Village Characteristics



Note: The figure depicts the results of a balance test on available observable baseline (standardized) characteristics of a village. In all cases the characteristic was used as a dependent variable in a baseline RDD. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. 95% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

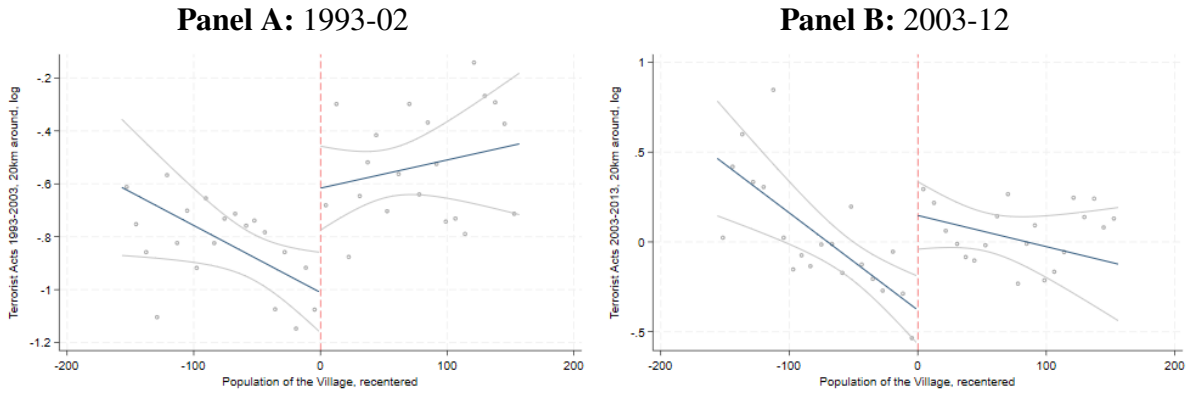
Figure 4: School Expansion and Terrorism. Pretreatment Placebos

Panel A: 1973-82. Pretreatment Phase **Panel B:** 1983-92. Construction Phase



Note: The figure depicts regression discontinuity plots for the analysis of the effect of school construction on the number terrorist attacks near the village in pretreatment period. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. The schooling reform was based on the 1981 census and implemented beginning in 1983. MSE-optimal bandwidth is selected. 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

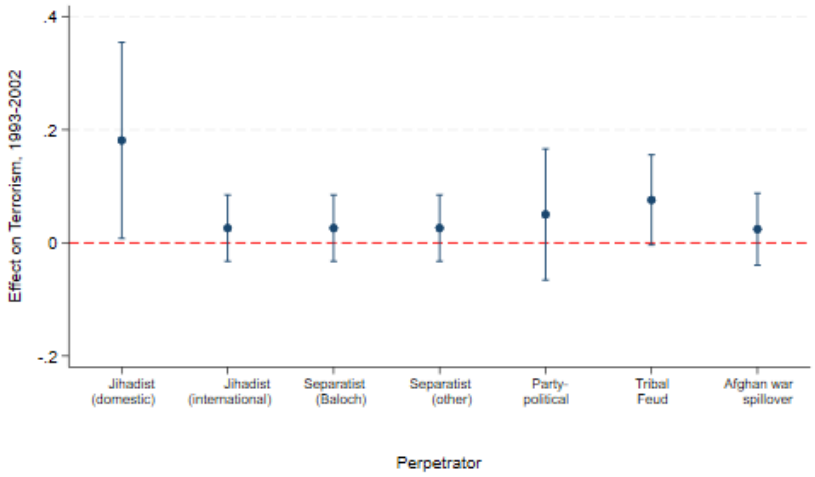
Figure 5: School Expansion and Terrorism.



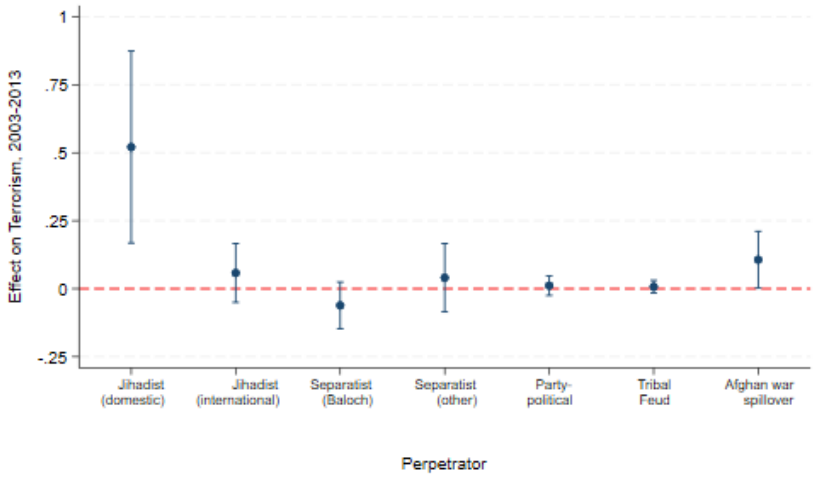
Note: The figure depicts regression discontinuity plots for the analysis of the effect of school construction on the number terrorist attacks near the village. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. The schooling reform was based on the 1981 census and implemented beginning in 1983. MSE-optimal bandwidth is selected. 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

Figure 6: Effect on Attacks by Perpetrator Type

Panel A:1993-02



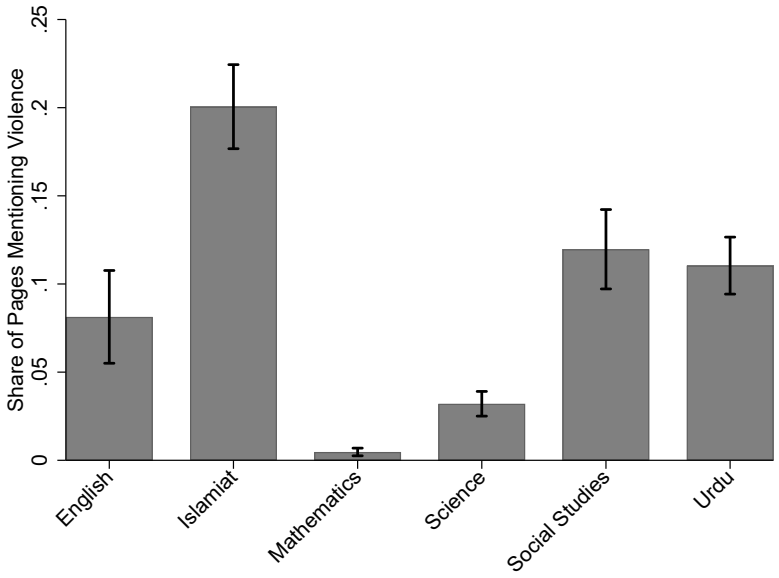
Panel B: 2003-12



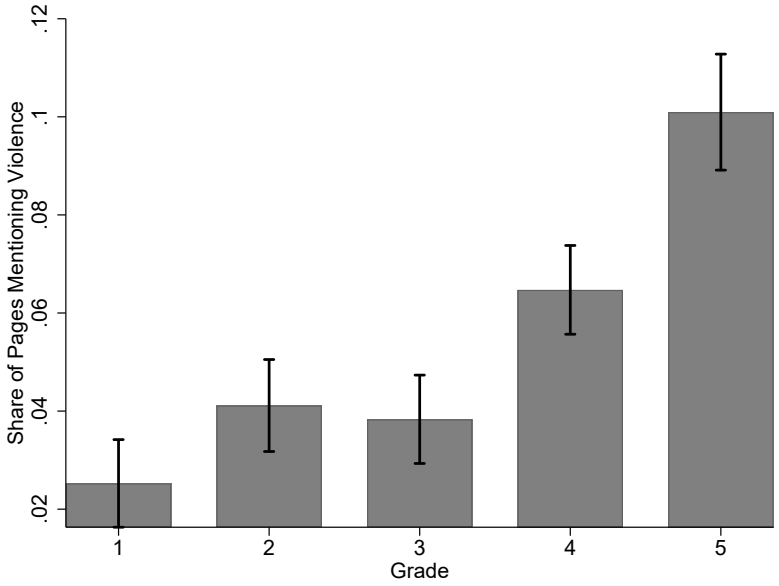
Note: the figure reports the analysis of the heterogeneity of the effect by perpetrator type. Groups that taken responsibility for terrorist attacks are clustered into broader classes and the baseline specification is estimated with the log number of attacks by the cluster as an outcome. Domestic jihadist group include the local terrorist organizations without international structure, such as Jamaat-ul-Ahrar; International Jihadists are terrorist organization with international chain of command, e.g. Al-Qaida; Balochi separatists are terrorist organization seeking the separation of Balochistan, such as BLA; Party-political are political structures engaging in terrorist attacks, such as MMA party, tribal are terrorist attacks related to the tribal feuds; Afghan war spillovers are the terrorist attacks organized by the participants of the war in Afghanistan, such as Taliban. For all regression MSE optimal bandwidth is selected and the coefficient is reported with 95% CI.

Figure 7: Violent Content in Textbooks

Panel A: By Subject

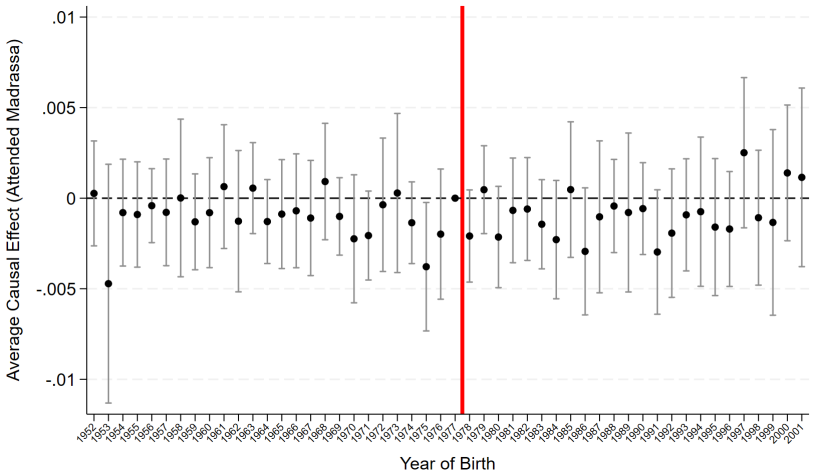


Panel B: By Grade



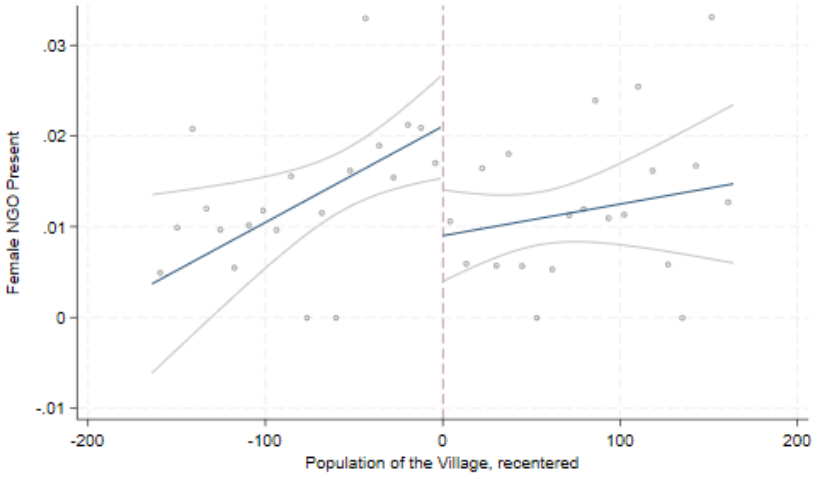
Note: the figure reports the result of textual analysis of textbooks context powered by a large language model. For each page Claude Sonnet 5 was asked to determine if the page had a reference to violent context. The reference to violent context was defined as "depiction or reference to physical force intended to hurt, kill, damage: war, battle, killing, beating, punishment, armed conflict, martyrdom involving violence. Sports/figurative "struggle" = metaphorical, set intensity low and category accordingly but you MAY still flag if clearly violent imagery." The mean of the variable is reported alongside with 95 CI.

Figure 8: Reform and Madrassa Attendance



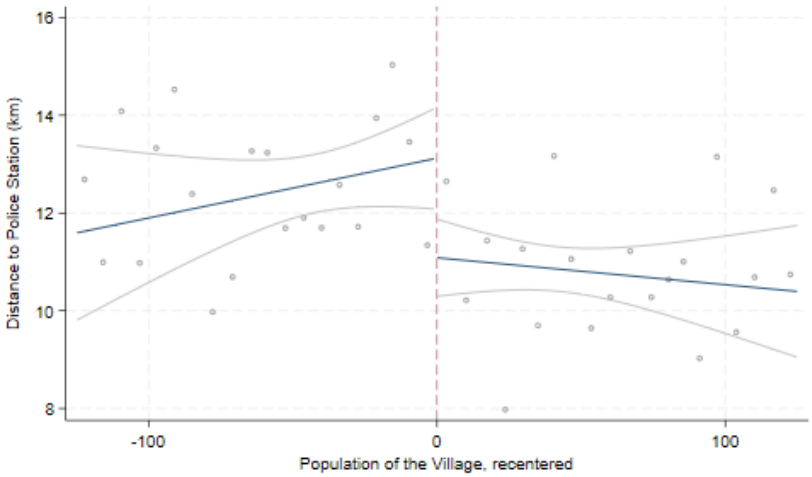
Note: The figure plots cohort-level event-study estimates using individual PSLM respondents. The design compares individuals born early versus late enough to be school-age after the school construction reform, across villages (relative to urban areas). Dependent variables is an indicator for attending a madrassa. Bars report 95% confidence intervals; standard errors are clustered at the district level. All specifications include district-by-urban, birth-year, and survey-year fixed effects.

Figure 9: School Construction and Female NGO



Note: The figure depicts regression discontinuity plots for the analysis of the effect of school construction on the female NGO activity in the village in 2020. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

Figure 10: School Construction and Police Stations



Note: The panel depicts regression discontinuity plots for the analysis of the effect of school construction on the distance to the closest police station in 2020. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

Table 1: Schooling Expansion, Education and Individual Level Skills

	Attended School	Can Read	Can Write	Can Count
	(1)	(2)	(3)	(4)
Rural × Born Post 1977	4.847***	4.949***	0.950***	3.105***
	(1.053)	(1.066)	(0.291)	(0.723)
District × Rural FE	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean Dep. Var.	51.832	55.101	97.869	77.445
Observations	471526	471526	259915	471526

Note: Robust standard errors clustered on district level in parentheses. The dependent variables are the indicators (multiplied by 100) of the respondent attending school, the ability to read a simple sentence, to write a simple sentence and the ability to solve a simple mathematical problem. The unit of observation is a respondent to PSLM survey in years 2007 2011 2013 2019. Treatment defined as person living in rural population and born after 1977 as the first cohort experienced the reform. Each specification controls for district by rural/urban fixed survey year and cohort fixed effects. Gender and age of the respondent are included as controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Schooling Expansion and Terrorism

	Log(Terrorist Attacks)			
	1973-1982	1983-1992	1993-2002	2003-2012
	(1)	(2)	(3)	(4)
School Constructed	0.051	0.169	0.394**	0.527**
	(0.062)	(0.147)	(0.165)	(0.208)
Local Polymoial, deg.	1	1	1	1
Mean Dep. Var.	-2.202	-1.035	-.673	-.097
Observations	13001	13001	13001	13001
Effective Observations	2497	2624	2535	2520

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variable is the logarithm of terrorist attacks in 20km proximity near the village in 10 year time period + 0.1. Outcomes for pretreatement 1973-83, and post treatment periods are presented The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. Specifications control for local polynomial of degree 1 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Schooling Expansion and Casualties

Panel A: Fatal Casualties				
	Log (# People Killed)			
	1973-1982	1983-1992	1993-2002	2003-2012
	(1)	(2)	(3)	(4)
School Constructed	0.049	0.254	0.557***	0.737***
	(0.056)	(0.184)	(0.211)	(0.259)
Mean Dep. Var.	-2.203	-1.022	-.243	-.01
Effective Observations	2435	2668	2548	2606
Panel B: Wounded				
	Log(# People Wounded)			
	1973-1982	1983-1992	1993-2002	2003-2012
School Constructed	0.053	0.193	0.493*	0.714***
	(0.058)	(0.232)	(0.254)	(0.276)
Local Polynomial, deg.	1	1	1	1
Mean Dep. Var.	-2.187	-.558	-.025	.235
Observations	13001	13001	13001	13001
Effective Observations	2435	2565	2505	2565

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variables are the logarithm of number of people killed in terrorist attacks in 20km proximity near the village in 10 year time period + 0.1 and the logarithm of number of people wounded in terrorist attacks in 20km proximity near the village in 10 year time period + 0.1. Outcomes for pretreatement 1973-83, and post treatment periods are presented The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. Specifications control for local polynomial of degree 1 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Schooling Expansion and Public Goods Provision

	Dist. to Police Station		Dist. to Health Center	
	(1)	(2)	(3)	(4)
School Constructed	-2.009*	-2.184*	0.075	0.173
	(1.118)	(1.162)	(1.107)	(1.184)
Local Polynomial, deg.	1	2	1	2
Mean Dep. Var.	9.651	9.651	9.223	9.223
Observations	33816	33816	33816	33816
Effective Observations	5531	10616	5933	10973

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variables are the distance to police office and rural health center per 2020 Mouza census in 2020. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. Specifications control for local polynomial of degree up to 2 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Schooling Expansion and NGOs

	Female NGO Present		Citizen Board NGO Present	
	(1)	(2)	(3)	(4)
School Constructed	-0.012**	-0.014**	0.006	0.009
	(0.006)	(0.007)	(0.007)	(0.007)
Local Polynomial, deg.	1	2	1	2
Mean Dep. Var.	.017	.017	.008	.008
Observations	33816	33816	33816	33816
Effective Observations	7247	12838	6214	11708

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variable is the indicator of presence of female and citizen board NGOs in the village per 2020 Mouza Census. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. Specifications control for local polynomial of degree up to 2 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Abou Zahab, Mariam and Olivier Roy, *Islamist Networks: The Afghan-Pakistan Connection* CERI Series in Comparative Politics and International Studies, New York: Columbia University Press, 2004. Translated from the French by John King.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of the Nazis in Prewar Germany," *Quarterly Journal of Economics*, 2015, *130* (4), 1885–1939.

Akresh, Richard, Daniel Halim, and Marieke Kleemans, "Long-term and intergenerational effects of education: Evidence from school construction in Indonesia," *The Economic Journal*, 2023, *133* (650), 582–612.

Alesina, Alberto, Paola Giuliano, and Bryony Reich, "Nation-Building and Education," *The Economic Journal*, 2021, *131* (638), 2273–2303.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "A Dime a Day: The Possibilities and Limits of Private Schooling in Pakistan," *Comparative Education Review*, 2008, *52* (3), 329–355.

_ , _ , _ , and Tristan Zajonc, "Religious School Enrollment in Pakistan: A Look at the Data," *Comparative Education Review*, 2006, *50* (3), 446–477.

Angrist, Joshua D. and Adriana D. Kugler, "Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia," *Review of Economics and Statistics*, 2008, *90* (2), 191–215.

Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena, "Bride Price and Female Education," *Journal of Political Economy*, 2020, *128* (2), 591–641.

Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo, "Nation-Building Through Compulsory Schooling During the Age of Mass Migration," *The Economic Journal*, 2019, *129* (617), 62–109.

Bano, Masooda, *The Rational Believer: Choices and Decisions in the Madrasas of Pakistan*, Ithaca, NY: Cornell University Press, 2012.

Bari, Faisal, Nadia Ejaz, and Adeel Faheem, "Role of NGOs in Primary Education in Pakistan: An Introduction," in "The Role of NGOs in Basic and Primary Education in Pakistan: NGO Pulse Report," Lahore, Pakistan: LUMS–McGill Social Enterprise Development Programme, Lahore University of Management Sciences, 2005, pp. 1–13.

Bazzi, Samuel and Christopher Blattman, "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 1–38.

_ , **Gabriel Koehler-Derrick, and Benjamin Marx**, "The Institutional Foundations of Religious Politics: Evidence from Indonesia," *The Quarterly Journal of Economics*, 2020, *135* (2), 845–911.

_ , **Masyhur Hilmy, and Benjamin Marx**, "Religion, Education, and the State," *Review of Economic Studies*, 2025.

Bengali, Kaiser, "History of Educational Policy Making and Planning in Pakistan," Working Paper Series 40, Sustainable Development Policy Institute, Islamabad, Pakistan 1999.

Benzer, Tolga and Janne Tukiainen, "Preaching to the Future: Religious Schools, Youth Organizations, and the Rise of Political Islam in Türkiye," Discussion Paper 171, Aboa Centre for Economics, Turku 2025. SSRN working paper 5394953.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig, "This Mine Is Mine! How Minerals Fuel Conflicts in Africa," *American Economic Review*, 2017, *107* (6), 1564–1610.

Berrebi, Claude, "Evidence about the Link Between Education, Poverty and Terrorism among Palestinians," *Peace Economics, Peace Science and Public Policy*, 2007, *13* (1), 18–53.

Besley, Timothy and Torsten Persson, *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton University Press, 2011.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs," Papers, arXiv.org 2018.

_ , _ , **and Rocio Titiunik**, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Campante, Filipe R. and Davin Chor, "Why Was the Arab World Poised for Revolution? Schooling, Economic Opportunities, and the Arab Spring," *Journal of Economic Perspectives*, 2012, *26* (2), 167–88.

Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Curriculum and Ideology," *Journal of Political Economy*, 2017, *125* (2), 338–392.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Center for Research and Security Studies, "Curriculum of Hate: Pakistan Focused. Independent Questions. Independent Answers," Report, Center for Research and Security Studies, Islamabad 2008.

Chen, Jiafeng and Jonathan Roth, "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 2024, *139* (2), 891–936.

Clark, Damon and Heather Royer, "The Effect of Education on Adult Mortality and Health: Evidence from Britain," *American Economic Review*, 2013, *103* (6), 2087–2120.

Crost, Benjamin, Joseph Felter, and Patrick Johnston, "Aid under Fire: Development Projects and Civil Conflict," *American Economic Review*, 2014, *104* (6), 1833–1856.

Cunha, Flavio and James J. Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

_ , _ , and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

de Mesquita, Ethan Bueno, "The Quality of Terror," *American Journal of Political Science*, 2005, *49* (3), 515–530.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

Dhar, Diva, Tarun Jain, and Seema Jayachandran, "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India," *American Economic Review*, 2022, *112* (3), 899–927.

Dube, Oeindrila and Juan F. Vargas, "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies*, 2013, *80* (4), 1384–1421.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 2001, *91* (4), 795–813.

_ , "The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia," *Journal of Development Economics*, 2004, *74* (1), 163–197.

Eggers, Andrew C., Ronny Freier, Veronica Grembi, and Tommaso Nannicini, "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions," *American Journal of Political Science*, 2018, *62* (1), 210–229.

Fetzer, Thiemo, "Did Austerity Cause Brexit?," *American Economic Review*, 2019, *109* (11), 3849–3886.

Flynn, Michael, "Michael Flynn once said most education in the Middle East is done in madrasas that are 'hate factories'," CNN 2016. Quote: "Most education in the Middle East is done in madrasas that are 'hate factories'.".

Fouka, Vasiliki, "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I," *Review of Economic Studies*, 2020, *87* (1), 204–239.

Giglio, Louis, Wilfrid Schroeder, and Christopher O. Justice, "The Collection 6 MODIS Active Fire Detection Algorithm and Fire Products," *Remote Sensing of Environment*, 2016, *178*, 31–41.

Haqqani, Husain, *Pakistan: Between Mosque and Military*, Washington, DC: Carnegie Endowment for International Peace, 2005.

Harari, Mariaflavia and Eliana La Ferrara, "Conflict, Climate, and Cells: A Disaggregated Analysis," *The Review of Economics and Statistics*, 2018, *100* (4), 594–608.

HBond, "Ghost Schools in Pakistan: A Crisis of Education and Accountability," https://hbond.org/ghost-schools-in-pakistan-a-crisis-of-education-and-accountability/ January 2025. Accessed: 2026-06-13.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), 411–482.

Imbens, Guido W. and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 2008, *142* (2), 615–635. The regression discontinuity design: Theory and applications.

Jia, Ruixue, "Weather Shocks, Sweet Potatoes and Peasant Revolts in Historical China," *Economic Journal*, 2014, *124* (575), 92–118.

Kaufman, Yoram J., Christopher O. Justice, Luke P. Flynn, Jackie D. Kendall, Elaine M. Prins, Louis Giglio, Darold E. Ward, W. Paul Menzel, and Alberto W. Setzer, "Potential Global Fire Monitoring from EOS-MODIS," *Journal of Geophysical Research: Atmospheres*, 1998, *103* (D24), 32215–32238.

Krueger, Alan B., *What Makes a Terrorist: Economics and the Roots of Terrorism*, Princeton, NJ: Princeton University Press, 2007.

_ **and Jitka Malečková**, "Education, Poverty and Terrorism: Is There a Causal Connection?," *Journal of Economic Perspectives*, 2003, *17* (4), 119–144.

LaFree, Gary and Laura Dugan, "Introducing the Global Terrorism Database," *Terrorism and Political Violence*, 2007, *19* (2), 181–204.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman, "Life Cycle Wage Growth across Countries," *Journal of Political Economy*, 2018, *126* (2), 797–849.

Martinez-Bravo, Monica and Andreas Stegmann, "In Vaccines We Trust? The Effects of the CIA's Vaccine Ruse on Immunization in Pakistan," *Journal of the European Economic Association*, 2022, *20* (1), 150–186.

Mazumder, Bhashkar, Maria Fernanda Rosales-Rueda, and Margaret Triyana, "Social Interventions, Health, and Well-Being: The Long-Term and Intergenerational Effects of a School Construction Program," *Journal of Human Resources*, 2023, *58* (4), 1097–1140.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714. The regression discontinuity design: Theory and applications.

McGuirk, Eoin and Marshall Burke, "The Economic Origins of Conflict in Africa," *Journal of Political Economy*, 2020, *128* (10), 3940–3997.

Mehmood, Sultan, Shaheen Naseer, and Daniel L. Chen, "Transmitting Rights: Effective Cooperation, Inter-gender Contact, and Student Achievement," *American Economic Journal: Economic Policy*, 2025, *17* (3), 107–130.

Meyersson, Erik, "Islamic Rule and the Empowerment of the Poor and Pious," *Econometrica*, 2014, *82* (1), 229–269.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112* (4), 725–753.

Mujahid-Mukhtar, Eshya, "Basic Education in Pakistan: Progress on EFA Target Dimensions in the 1990s," Background Study, UNESCO Islamabad, Islamabad September 1999.

Nasr, Vali R., "International Politics, Domestic Imperatives, and Identity Mobilization: Sectarianism in Pakistan, 1979–1998," *Comparative Politics*, 2000, *32* (2), 171–190.

National Consortium for the Study of Terrorism and Responses to Terrorism, "Global Terrorism Database™ (GTD)," 2020. Accessed: 2025-10-04.

National Planning Board, Government of Pakistan, *The First Five Year Plan, 1955–60*, Government of Pakistan, 1957. December 1957.

Nayyar, A. H. and Ahmad Salim, eds, "The Subtle Subversion: The State of Curricula and Textbooks in Pakistan," Report, Sustainable Development Policy Institute, Islamabad 2003.

Nunn, Nathan and Nancy Qian, "US Food Aid and Civil Conflict," *American Economic Review*, 2014, *104* (6), 1630–1666.

Pakistan Bureau of Statistics, "Mouza Census-2020," Technical Report, Pakistan Bureau of Statistics 2020. Accessed: 2026-06-13.

_ , "Pakistan Social and Living Standards Measurement Survey (PSLM) 2019-20," 2021. National / Provincial / District report, Islamabad, July 2021.

Pape, Robert A. and James K. Feldman, *Cutting the Fuse: The Explosion of Global Suicide Terrorism and How to Stop It*, Chicago: University of Chicago Press, 2010.

Planning Commission, Government of Pakistan, *The Sixth Five Year Plan, 1983–88*, Planning Commission, Government of Pakistan, 1983.

_ , *The Seventh Five Year Plan, 1988–93*, Islamabad: Government of Pakistan, 1988.

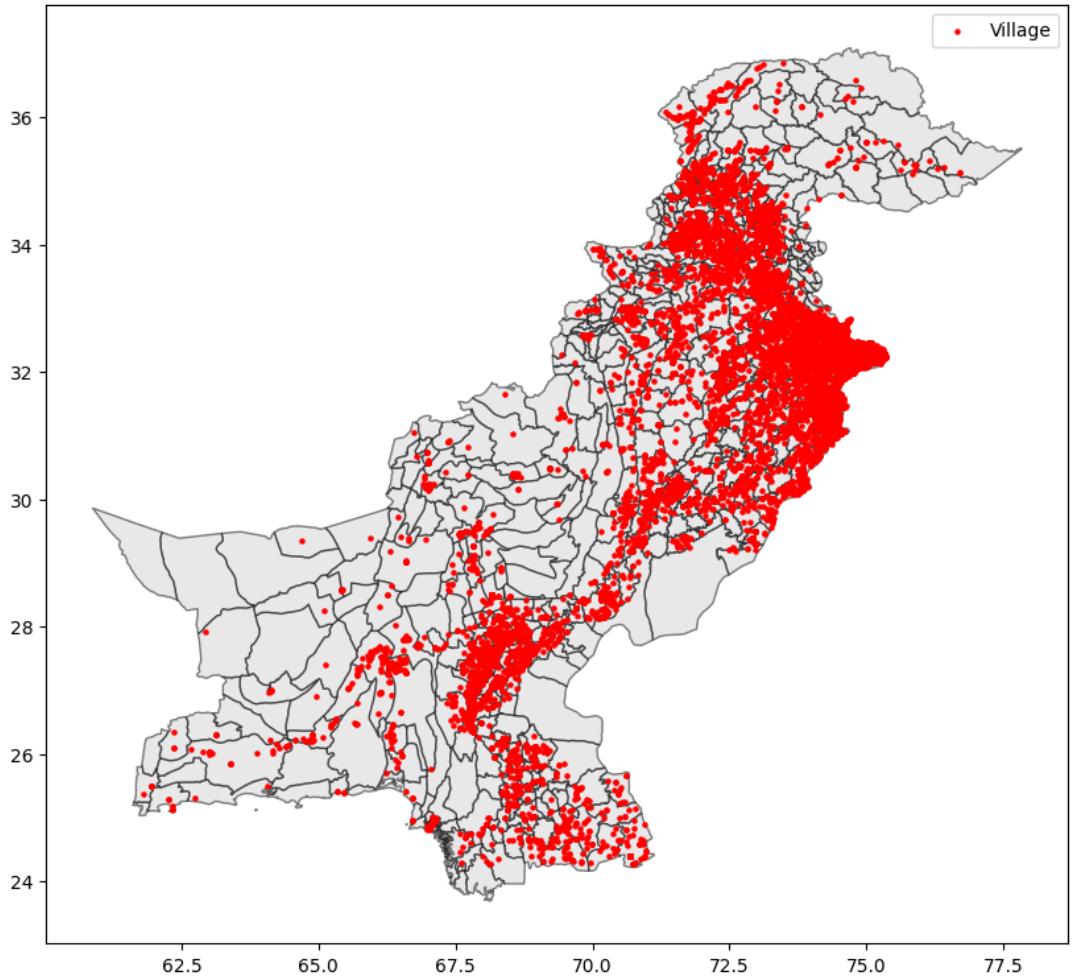
Saleem, Hina, "Bridging the Gap: A Holistic Education Policy to Foster Opportunities for Girls in Rural Pakistan," 2022.

UNESCO International Institute for Educational Planning, "Study on Primary School Curriculum and Textbook in Pakistan," IIEP Research Report, UNESCO, Paris 2004.

Voigtländer, Nico and Hans-Joachim Voth, "Nazi Indoctrination and Anti-Semitic Beliefs in Germany," *Proceedings of the National Academy of Sciences*, 2015, *112* (26), 7931–7936.

9. Appendix A. Additional Tables and Figures

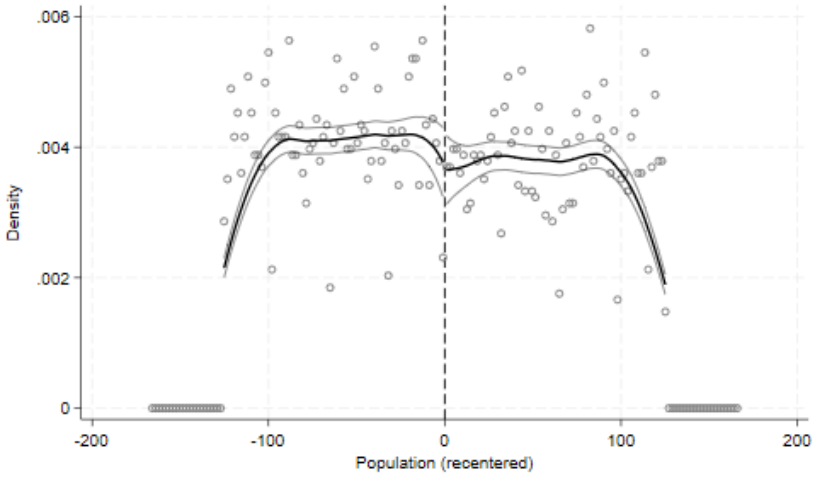
Figure A1: Map of Villages in Pakistan



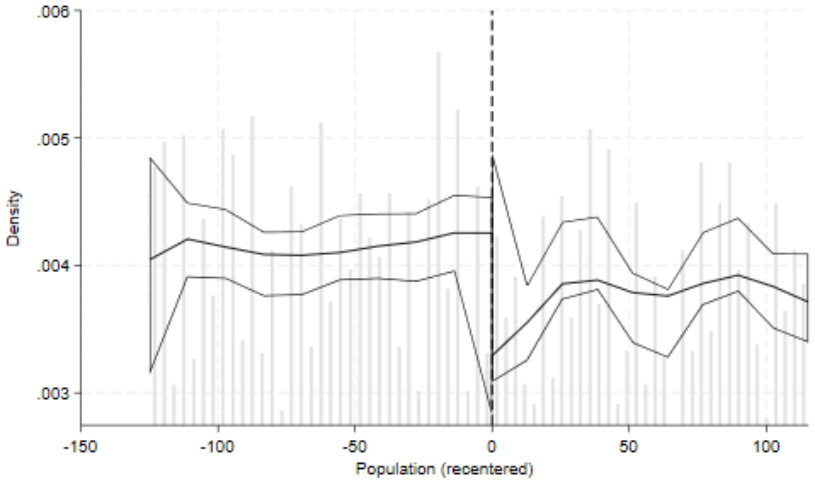
Note: The figure reports the geographical distribution of villages across Pakistan. The villages data is from 1981 Census and geolocated using Google maps and Nominatim API of Open Street Map.

Figure A2: Continuity at the Cutoff for 2020 Sample

Panel A: McCrary (2008) Test



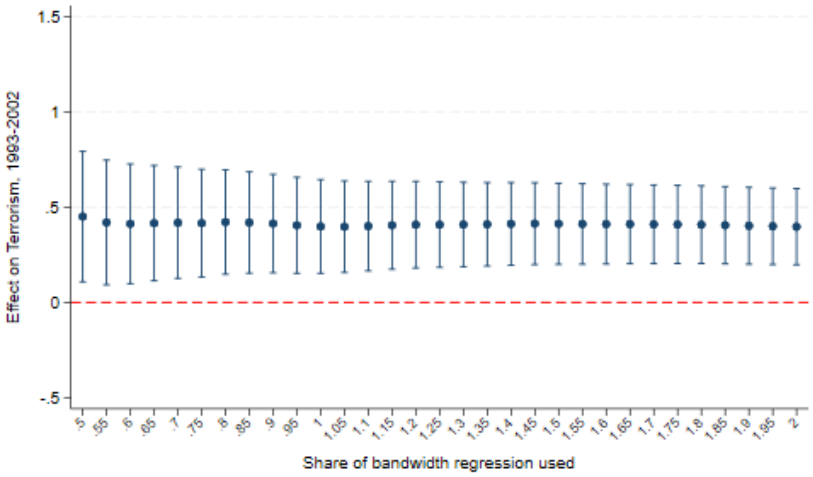
Panel B: Cattaneo et al. (2020) Test



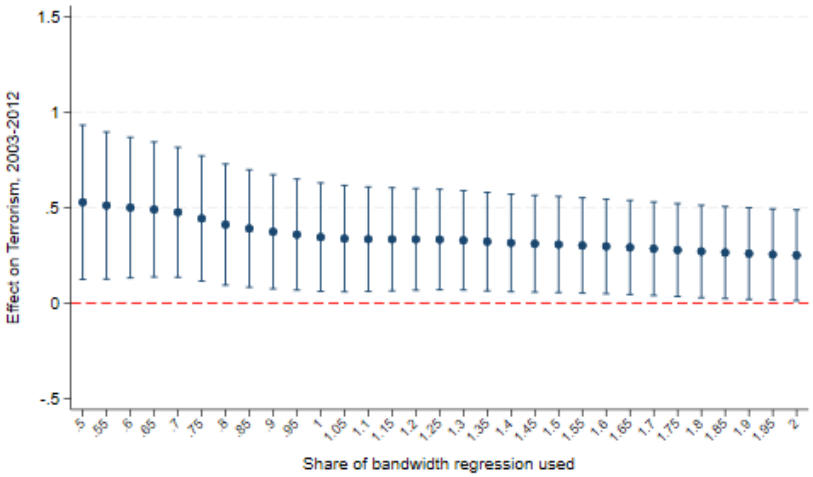
Note: The figure reports the distribution of the village population within the MSE-optimal bandwidth and two density tests. The density is estimated on the both side of the threshold to check the continuity, p-value of McCrary test 0.85 and for Cattaneo et al. is 0.694

Figure A3: Bandwidth Sensitivity: Terrorist Attacks

Panel A: 1993-02 Attacks



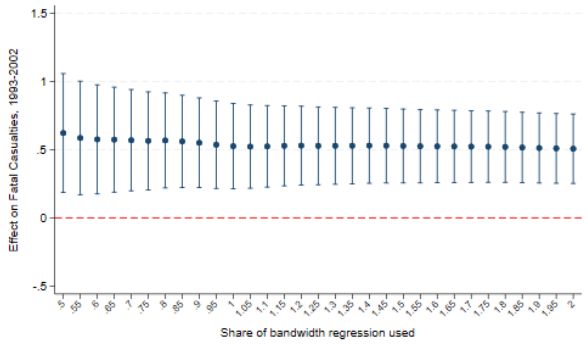
Panel B: 2003-12 Attacks



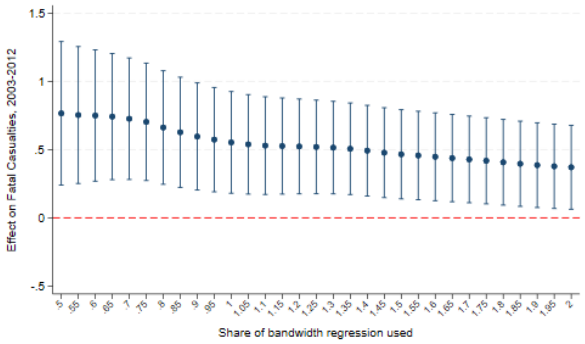
Note: The figure reports the sensitivity of the main results to the choice of bandwidth. The baseline regression with local polynomial of degree one is estimated for the sample restricted as a share of the baseline MSE-optimal bandwidth. This figure reports the test for 2 outcomes: log number of terrorist attacks in 1993-02, 2003-12 +0.1. Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator and 95% CI are reported.

Figure A4: Bandwidth Sensitivity: Casualties

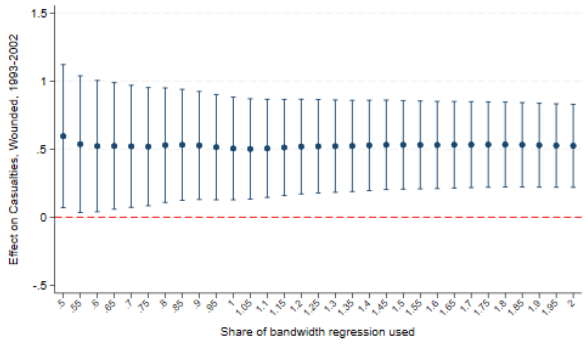
Panel A: Fatal Casualties 1993-02



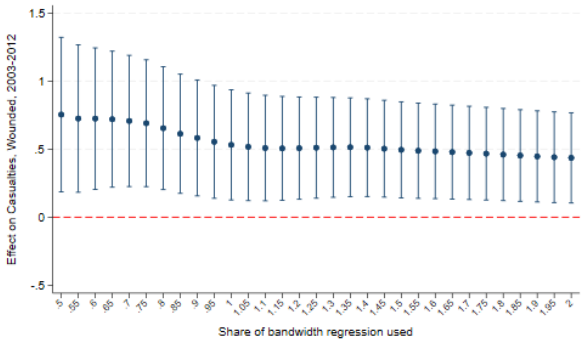
Panel B: Fatal Casualties 2003-12



Panel C: Casualties, Wounded, 1993-02



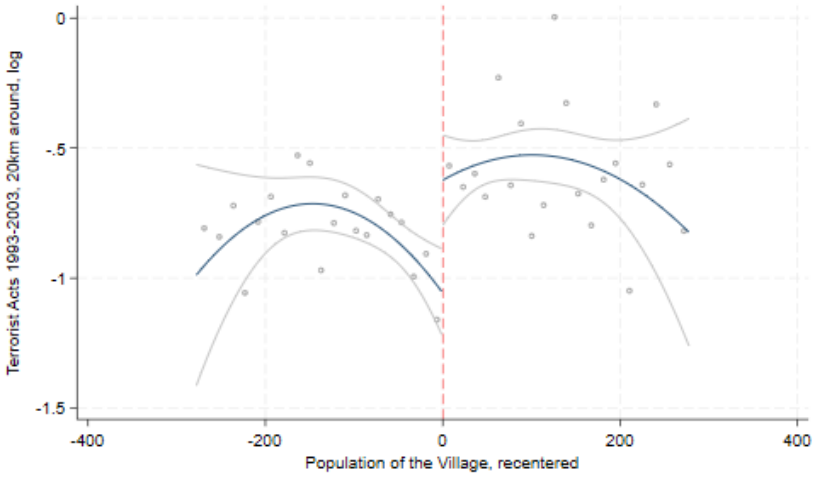
Panel D: Casualties, Wounded, 2003-12



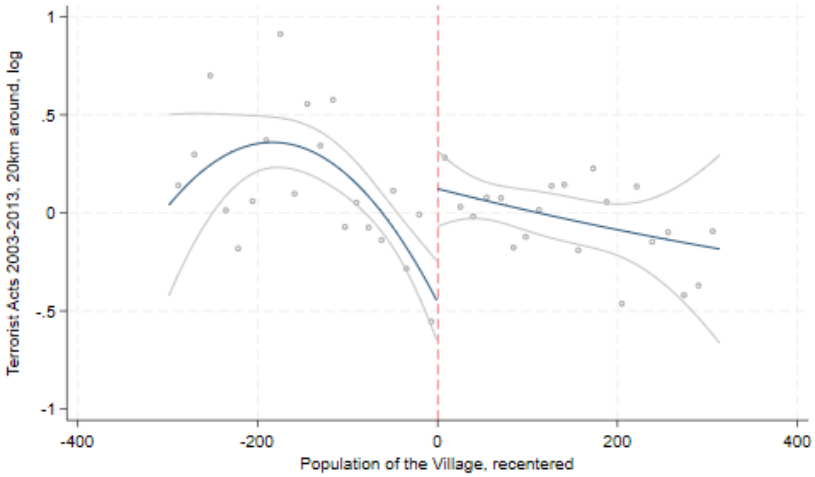
Note: The figure reports the sensitivity of the main results to the choice of bandwidth. The baseline regression with local polynomial of degree one is estimated for the sample restricted as a share of the baseline MSE-optimal bandwidth. This figure reports the test for 4 outcomes: log number of killed and wounded in 1993-02 and 2003-12 + 0.1. Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator and 95% CI are reported.

Figure A5: Sensitivity of Terrorist Attacks Result to Polynomial Degree

Panel A: 1993-02 Attacks



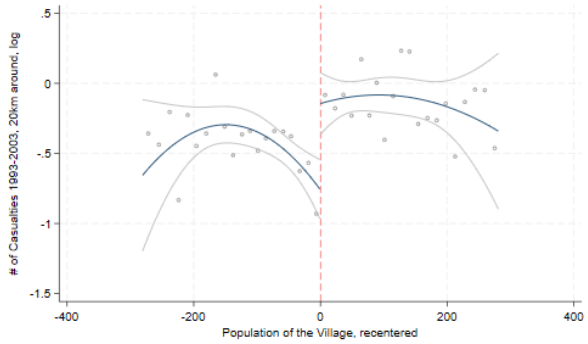
Panel B: 2003-12 Attacks



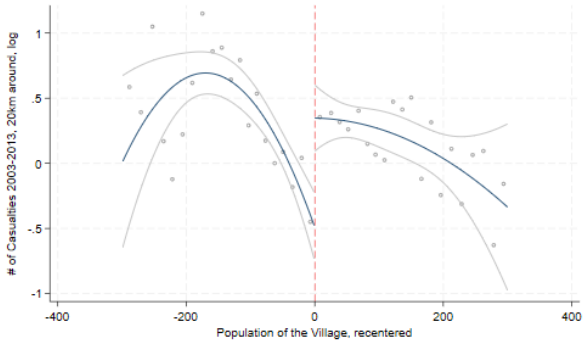
Note: The figure reports the sensitivity of the main results to the choice of polynomial degree. The baseline regression with local polynomial of degree two is estimated for the sample of the baseline MSE-optimal bandwidth. This figure reports the test for 2 outcomes: log number of terrorist attacks in 1993-02, 2003-12 +0.1. Standard errors clustered on a district level and 90% CI are reported.

Figure A6: Sensitivity of Casualties Result to Polynomial Degree

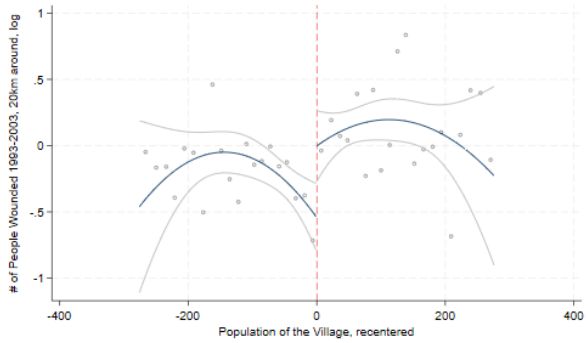
Panel A: Fatal Casualties 1993-02



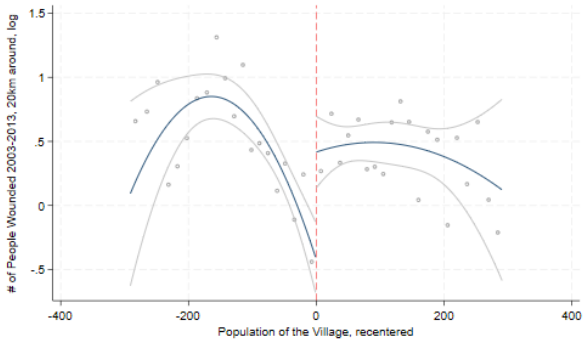
Panel B: Fatal Casualties 2003-12



Panel C: Casualties, Wounded, 1993-02



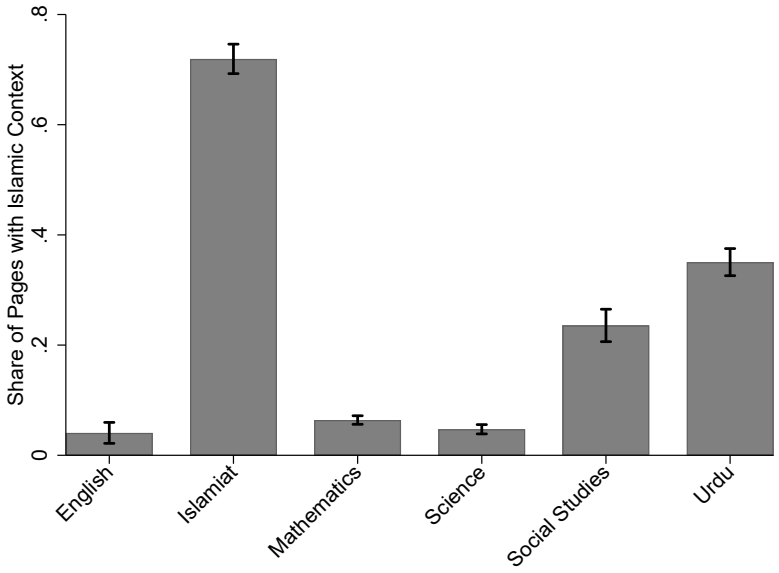
Panel D: Casualties, Wounded, 2003-12



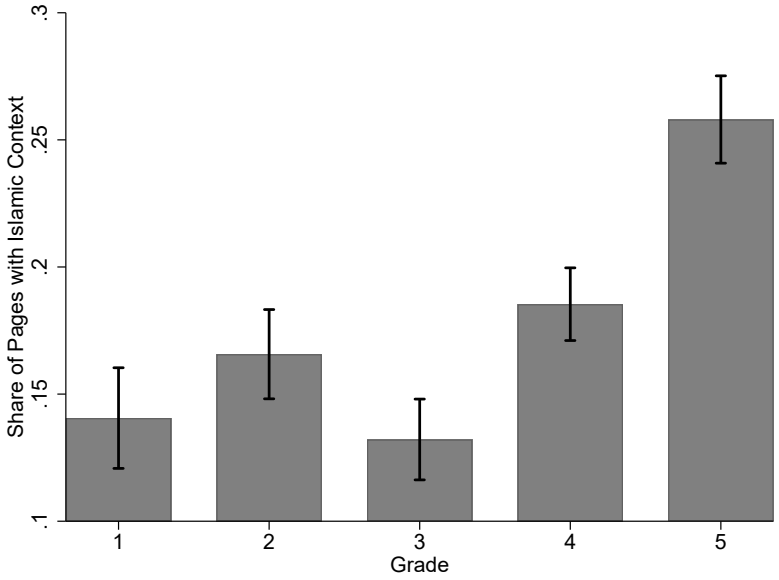
Note: The figure reports the sensitivity of the main results to the choice of polynomial degree. The baseline regression with local polynomial of degree two is estimated for the sample of the baseline MSE-optimal bandwidth. This figure reports the test for 4 outcomes: log number of killed and wounded in 1993-02 and 2003-12 + 0.1. 90% CI are reported.

Figure A7: Islamic Content in Textbooks

Panel A: By Subject

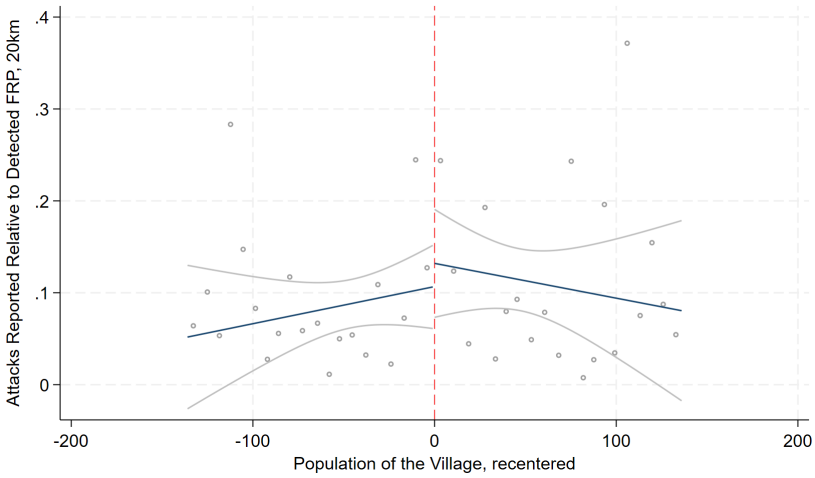


Panel B: By Grade



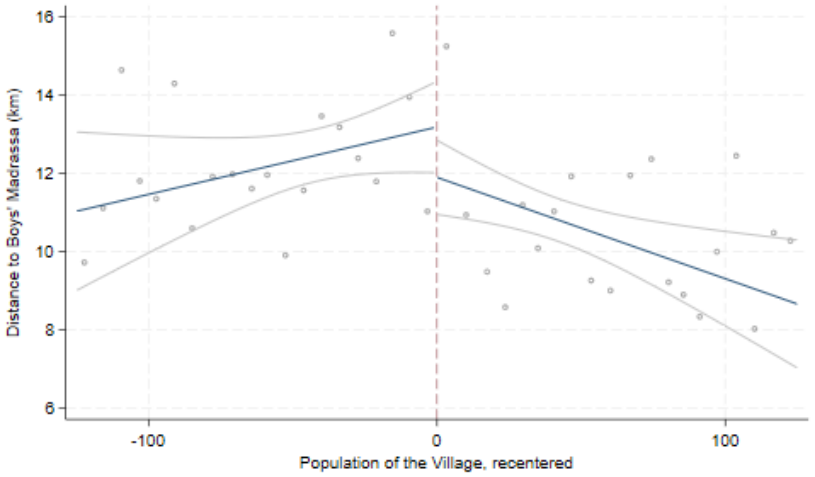
Note: the figure reports the result of textual analysis of textbooks context powered by a large language model. For each page Claude Sonnet 5 was asked to determine if the page had a reference to Islamic context. The reference to Islamic context was defined as "explicit reference to Islam, Muslims, the Prophet, Quran, Islamic practice/figures/history. Generic mention of God in a non-Islamic frame is NOT Islam unless context is Islamic." The mean of the variable is reported alongside with 95 CI.

Figure A8: School Construction and Attacks Reporting



Note: The figure depicts regression discontinuity plot for the analysis of the effect of school construction on the reporting of the terrorist attacks. The dependent variable is the ratio of the number of attacks and total Fire Radiative Power captured by NASA satellites in 2003-2012. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

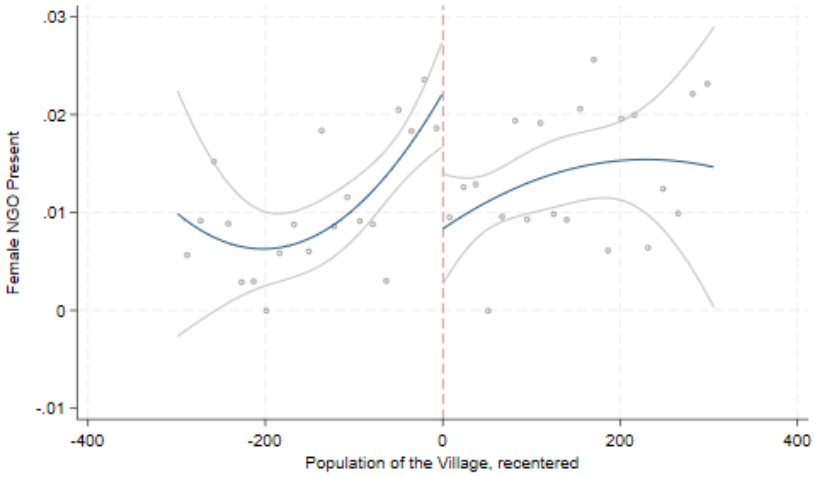
Figure A9: School Construction and Madrassa Expansion



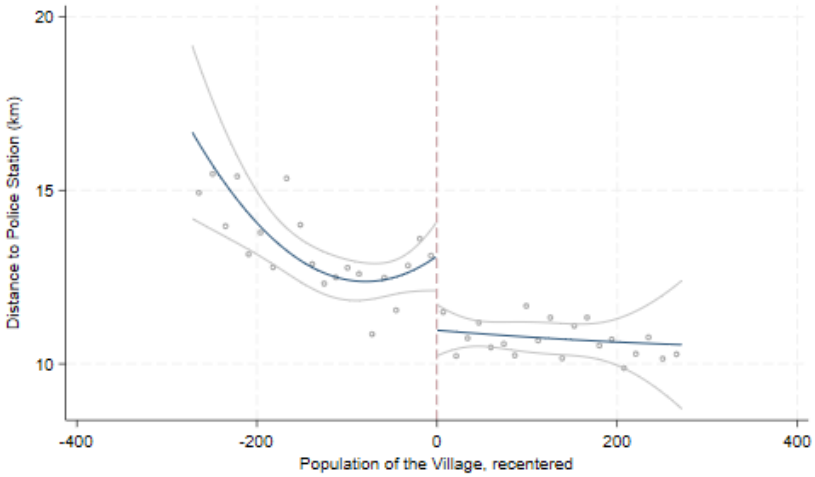
Note: The figure depicts regression discontinuity plots for the analysis of the effect of school construction on the distance to the closest religious school in 2020. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

Figure A10: Sensitivity of Distance to Police and NGO Presence Results to Polynomial Degree

Panel A: Female NGO Presence



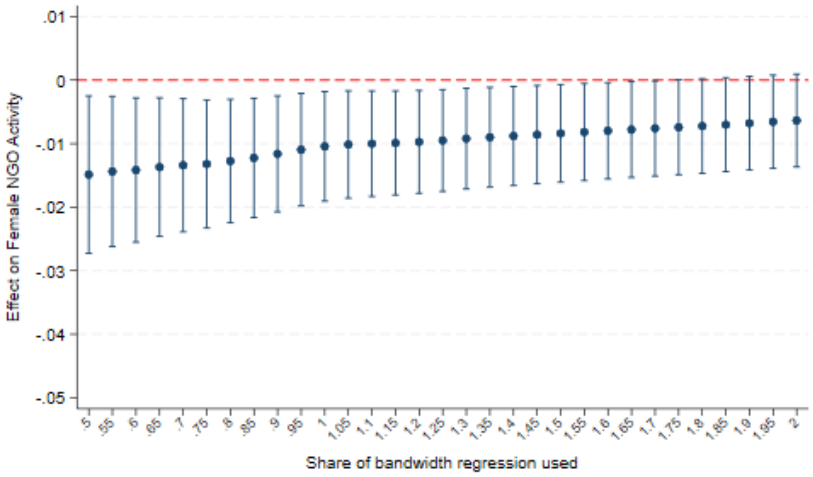
Panel B: Distance to Police Station



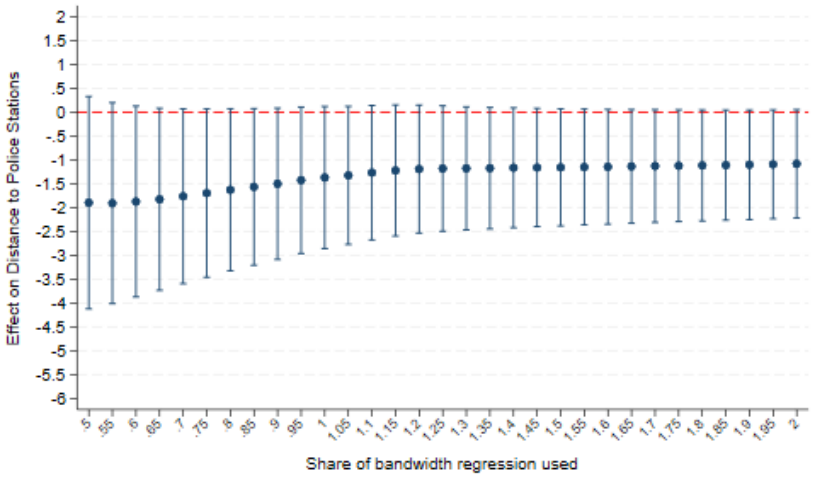
Note: The figure reports the sensitivity of the main results to the choice of polynomial degree. The baseline regression with local polynomial of degree two is estimated for the sample of the baseline MSE-optimal bandwidth. This figure reports the test for 2 outcomes: presence of female NGO and distance to police station as per 2020 Mouza Census. 90% CI are reported.

Figure A11: Bandwidth Sensitivity: Terrorist Attacks

Panel A: Female NGO Presence



Panel B: Distance to Police Station



Note: The figure reports the sensitivity of the main results to the choice of bandwidth. The baseline regression with local polynomial of degree one is estimated for the sample restricted as a share of the baseline MSE-optimal bandwidth. This figure reports the test for 2 outcomes: presence of female NGO and distance to police station as per 2020 Mouza Census. Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator and 95% CI are reported.

Table A1: Summary Statistics of Main Variables

Panel A: Villages and Terrorism						
	Mean	**Median**	**SD**	**Min**	**Max**	**N obs**
Population, 1981	1413.9	927	1695.8	1	28841	13001
# of Terrorist Attacks, 1973-82 log	-2.202	-2.303	0.626	-2.303	2.208	13001
# of Terrorist Attacks, 1983-92 log	-1.035	-2.303	1.737	-2.303	4.490	13001
# of Terrorist Attacks, 1993-02 log	-0.673	-2.303	1.958	-2.303	4.278	13001
# of Terrorist Attacks, 2003-12 log	-0.0971	0.0953	2.313	-2.303	6.306	13001
# of People Killed in Terrorist Attacks, 1973-82 log	-2.203	-2.303	0.651	-2.303	3.182	13001
# of People Killed in Terrorist Attacks, 1983-92 log	-1.022	-2.303	2.118	-2.303	5.038	13001
# of People Killed in Terrorist Attacks, 1993-02 log	-0.243	-2.303	2.506	-2.303	5.070	13001
# of People Killed in Terrorist Attacks, 2003-12 log	-0.0100	-2.303	2.911	-2.303	6.976	13001
# of People Wounded in Terrorist Attacks, 1973-82 log	-2.187	-2.303	0.756	-2.303	4.096	13001
# of People Wounded in Terrorist Attacks, 1983-92 log	-0.558	-2.303	2.684	-2.303	6.577	13001
# of People Wounded in Terrorist Attacks, 1993-02 log	-0.0251	-2.303	2.934	-2.303	6.325	13001
# of People Wounded in Terrorist Attacks, 2003-13 log	0.235	-2.303	3.188	-2.303	7.897	13001
Panel B: Villages and 2020 Outcomes						
	Mean	**Median**	**SD**	**Min**	**Max**	**N obs**
Population, 1981	1342.7	852	1599.3	1	31075	33816
Distance to Police Station, 2020	9.651	7	13.96	0	830	33816
Distance to Boys Madrassa, 2020	9.142	5	14.83	0	301	33816
Distance to Girls Madrassa, 2020	11.99	6	20.31	0	515	33816
Female NGO Present, 2020	0.0171	0	0.130	0	1	33816
Distance to Rural Health Center, 2020	9.223	5	15.44	0	754	33816
Citizen Board NGO Present, 2020	0.00840	0	0.0913	0	1	33816
Panel C: Survey Outcomes						
	Mean	**Median**	**SD**	**Min**	**Max**	**N obs**
Educated	0.545	1	0.498	0	1	706273
Lives in Rural Area	0.655	1	0.475	0	1	706273
Birth Year	1984.0	1987	13.46	1952	2002	706273
Can Read	0.565	1	0.496	0	1	687766
Can Count	0.817	1	0.387	0	1	687769
Can Write	0.982	1	0.136	0	2	388673
Female	0.502	1	0.500	0	1	706273
Age	32.13	30	14.21	3	80	706273

Notes: This table reports mean, median, standard deviation for main in our analysis. Panel A covers dataset of terrorism and villages, Panel B reports summary statistics for merged 1981 village and 2020 mouza census dataset, while Panel C covers the PIHS survey. The unit of analysis is village in Panels A and B and PIHS survey response in Panel C. Attacks and were merged using 20km buffer.

Table A2: Robustness to Functional Form

	Panel A: 1993-02				
	Terrorist Attacks, 1993-02				
	log(X + 1)	log(X + 0.01)	ihs(X)	PPML	levels
	(1)	(2)	(3)	(4)	(5)
main					
School Constructed	0.168**	0.931***	0.209**	0.314	0.771
	(0.083)	(0.356)	(0.103)	(0.206)	(0.557)
Mean Dep. Var.	.733	-3.24	.921	3.314	3.314
Effective Observations	2505	2548	2505	2535	2535
	Panel B: 2003-12				
	Terrorist Attacks, 2003-12				
	log(X + 1)	log(X + 0.01)	ihs(X)	PPML	levels
main					
School Constructed	0.350***	0.869**	0.413***	0.496*	5.646*
	(0.133)	(0.383)	(0.155)	(0.283)	(3.354)
Local Polynomial, deg.	1	1	1	1	1
Mean Dep. Var.	1.094	-2.195	1.336	12.79	12.79
Observations	13001	13001	13001	13001	13001
Effective Observations	2382	2681	2394	2364	2364

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variables are the logarithm + 1, logarithm + 0.01, inverse hyperbolic sine, levels of number of terrorist attacks in 20km proximity near the village in 10 year time period for post treatment periods 1993-2002, 2003-2012. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected for poisson regression the bandwidth is selected based on the linear counterpart. Specifications control for local polynomial of degree 1 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Schooling Expansion and Growth of Settlement

	Is Urban		Is Partly Urban	
	(1)	(2)	(3)	(4)
School Constructed	0.011	0.012	-0.001	-0.002
	(0.008)	(0.008)	(0.004)	(0.005)
Local Polynomial, deg.	1	2	1	2
Mean Dep. Var.	.028	.028	.026	.026
Observations	33816	33816	33816	33816
Effective Observations	5110	11348	6926	12470

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variables are the indicator of the village being urban or partly urban per 2020 Mouza census in 2020. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. Specifications control for local polynomial of degree 2 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Pakistan Primary School Curriculum: Time Allocation per Week by Class

Subject	**Class I**		**Class II**		**Class III**		**Class IV**		**Class V**	
	Time	%	Time	%	Time	%	Time	%	Time	%
Maths	6h 40m	25.0%	6h 40m	25.0%	6h 40m	25.0%	6h 40m	25.0%	6h 40m	25.0%
Urdu	6h 40m	25.0%	6h 40m	25.0%	6h 40m	25.0%	6h	22.5%	6h	22.5%
English	6h 40m	25.0%	6h 40m	25.0%	6h 40m	25.0%	3h 20m	12.5%	3h 20m	12.5%
Social Studies	–	–	–	–	–	–	3h 20m	12.5%	3h 20m	12.5%
Science / G. Knowledge	2h 40m	10.0%	2h 40m	10.0%	2h 40m	10.0%	3h 20m	12.5%	3h 20m	12.5%
Islamiyat	2h 40m	10.0%	2h 40m	10.0%	2h 40m	10.0%	2h 40m	10.0%	2h 40m	10.0%
Art	40m	2.5%	40m	2.5%	40m	2.5%	40m	2.5%	40m	2.5%
Games	40m	2.5%	40m	2.5%	40m	2.5%	40m	2.5%	40m	2.5%
Total	26 hrs 40 minutes per week (100%)									

Note: This table reports the curriculum of primary schools in Pakistan in 1980-90 as per UNESCO International Institute for Educational Planning (2004).

Table A5: Schooling Expansion and Religious Schooling

	Dist. to Boys Madrassa		Dist. to Girls Madrassa	
	(1)	(2)	(3)	(4)
School Constructed	-1.250	-1.559	-0.292	-0.248
	(1.265)	(1.331)	(1.499)	(1.594)
Local Polynomial, deg.	1	2	1	2
Mean Dep. Var.	9.142	9.142	11.995	11.995
Observations	33816	33816	33816	33816
Effective Observations	5405	10164	5933	11035

Note: Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator in parentheses. The dependent variables is distance to religious schools for boys and girls per 2020 Mouza census. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. MSE-optimal bandwidth is selected. Specifications control for local polynomial of degree 2 on the both side of the threshold. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

10. Appendix B. Dataset Construction

This appendix describes the construction process of our datasets. We start with the 1981 Census of Pakistan. Panel A of Figure B1 reports the sample of the SELECTED POPULATION STATISTICS OF INDIVIDUAL RURAL LOCALITIES table of the Census. As a first step we digitized this table across 62 district Census reports covering the whole country of Pakistan. The digitization process relied on the Mistral AI API, the exact prompt for the extraction can be found in Appendix C. The goal of the model was to return a json for every village in the census. After this procedure we ended up with approximately 40000 villages. The manual cross check of 500 random sampled villages revealed no mistakes in the data extraction comparing to the raw census data.

The census pages report two kinds of rows: individual localities and Police Circle or Tehsil subtotals, the latter marked with an asterisk. The model was instructed to flag the subtotals, and we drop them, so that the unit of observation is a single rural locality. We then convert all demographic fields to numeric values and check the internal consistency of the extraction: total population equals the sum of male and female population, literacy rates lie in $[0, 100]$, and hadbasti (locality) numbers are unique within a district. Population, the running variable of the design, is reported in every district report and is never imputed.

Village Coordinates. The 1981 Census reports no coordinates, so the villages have to be geolocated before they can be linked to anything spatial. We query the Google Maps Geocoding API with the string *"{village name}, {district}, Pakistan"*, restricted to Pakistan, and keep the first

returned result together with its coordinates, its formatted address, and its location type. Since the 1981 districts in Pakistan are different from the current, we used a crosswalk, that mapped possible modern districts to the 1981 ones. Google reports the location type as ROOFTOP, GEOMETRIC_CENTER, RANGE_INTERPOLATED, or APPROXIMATE, and this label is what allows us to separate genuine settlement points from coarser ones.

The APPROXIMATE returns are the problematic ones. Google resolves a village name it does not know to the center of the enclosing administrative unit rather than returning nothing, which places distinct villages on a single tehsil or district point. We therefore re-geocode every APPROXIMATE locality against the Nominatim API of OpenStreetMap, which reports the class and type of the object it matched. Results typed as village, hamlet, or isolated dwelling are settlement points; those typed as administrative or municipality are again unit centroids. The final coordinate of a village is the Google coordinate when Google returned a precise location type, and the OpenStreetMap coordinate otherwise. Villages for which neither source returns a settlement point are dropped, since we cannot assign them a location. This leaves 13001 villages with usable coordinates, of which 2231 lie below the 300 person threshold. Figure A1 maps them. While there might be some selection into geolocation in out sample, Figure B2 shows that there is no discontinuity near the population threshold in propensity of the coordinates being available, so this is unlikely to undermine our conclusions.

Education and Terrorist Attacks. We link the geolocated villages to the Global Terrorism Database, which records the date, location, perpetrator and casualties of attacks worldwide since 1970. We keep attacks in Pakistan with non-missing coordinates and compute the great circle distance between every attack and every village. An attack is assigned to a village if it occurred within 20 kilometers of it. The buffer is not a partition: an attack in an area where villages are dense is assigned to each village within the radius, and a village with no attack in its radius is assigned a count of zero rather than dropped. Figure B3 reports the robustness to the choice of this assignment threshold and across other possible options, the effect of interest remains stable. For the purpose of our analysis we drop the villages that were in the middle of Pakistan military operations and ended up in 99th percentile of the distribution. For each village we then count attacks, people killed, and people wounded within the buffer in each of the four ten year windows 1973–1982, 1983–1992, 1993–2002 and 2003–2012, and take logs of the counts after adding 0.1. Figure B4 reports the effect on the terrorist attacks by year, rather than decades and shows that the aggregation do not drive the results. We also count attacks separately by perpetrator cluster, assigning every group named in the database to one of seven clusters.

Matched Census Dataset. The 2020 Mouza Census records, for every rural mouza in Pakistan, the identification codes of the province, division, district, tehsil and mouza, and the answers to a questionnaire covering private and public amenities of a village. Panel B of Figure B1 reports a page of the questionnaire. Amenities that the mouza does not have are recorded as the distance in kilometers to the nearest one, which is the form our 2020 outcomes take. We recode the questionnaire item codes into named variables using the published form.

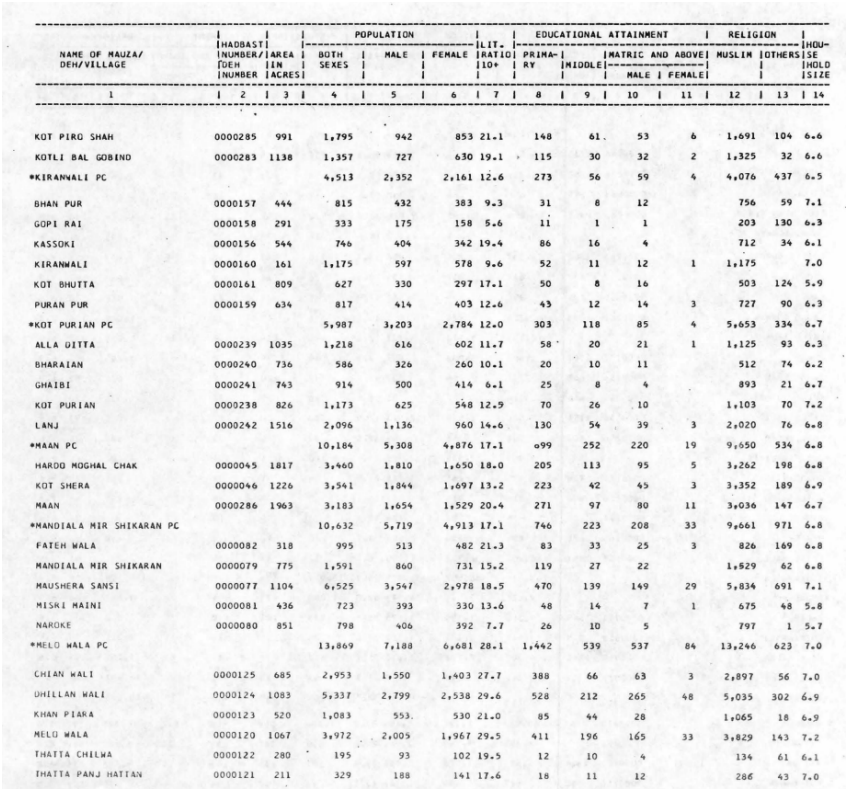
The two censuses share no common village identifier: hadbasti numbers were not carried over into the 2020 mouza codes. We therefore match on names within districts. District boundaries changed substantially over the four decades, so we first build a crosswalk that maps each of the 62 census districts of 1981 to the full set of its 2020 successor districts, so that, for instance, the 1981 district of Swat is searched against Swat, Shangla and Buner in 2020. Names are standardized on both sides before matching: we upper-case them, remove parenthetical qualifiers and punctuation, replace hyphens and slashes by spaces, collapse repeated whitespace, and correct the recurring OCR errors. The first pass accepts only equality of standardized names within the successor pool. The second pass takes the villages still unmatched and, for each, the best candidate in the same pool under the token sort ratio, accepting it if the ratio exceeds 85. The two passes together return 33816 matched villages, of which nearly 7000 lie below the 300 person threshold.

Individual-Level Survey. The individual-level dataset pools the rounds of the Pakistan Social and Living Standards Measurement survey. We keep respondents born between 1952 and 2001, and construct the indicators for school attendance and for the ability to read, write and count, together with the rural indicator, gender, age, district of residence and survey year that the specification uses.

Textbook Corpus. The curriculum evidence comes from 132 Urdu and English language primary school textbooks for Grades 1–5, roughly 10000 pages, recovered from the collection of one of the largest libraries in Pakistan. Each page is rendered as an image and sent once to Claude Sonnet 5, which returns a faithful English rendering of the page, translating it when the page is in Urdu, together with the page type, the subject and edition year when they are printed on the page, and the coding of Islamic and violent content: an indicator, a count of distinct references, a category, and short verbatim excerpts justifying each flag. The requirement to quote the text that justifies a flag is what makes the coding auditable. We drop covers, tables of contents and blank pages and keep content and exercise pages, so that the unit of analysis is a page of instructional material. Appendix C reports the exact prompt.

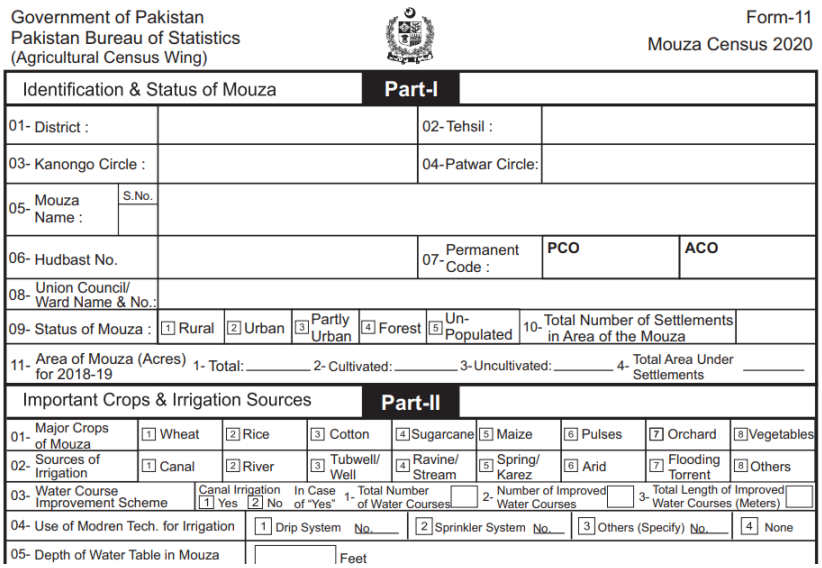
Figure B1: Sample Data Sources

Panel A: 1981 Census



NAME OF MAUZA/ DEH/VILLAGE	HADBAST NUMBER/ DEH NUMBER	AREA IN ACRES	POPULATION BOTH SEXES	MALE	FEMALE	LIT. RATIO 10+	EDUCATIONAL ATTAINMENT PRIMARY	MIDDLE	MATRIC AND ABOVE MALE	MATRIC AND ABOVE FEMALE	RELIGION MUSLIM	OTHERS	HOUSE HOLD SIZE
1	2	3	4	5	6	7	8	9	10	11	12	13	14
KOT PIRO SHAH	0000285	991	1,795	942	853	21.1	148	61	53	6	1,691	104	6.6
KOTLI BAL GOBIND	0000283	1138	1,357	727	630	19.1	115	30	32	2	1,325	32	6.6
*KIRANWALI PC			4,513	2,352	2,161	12.6	273	56	59	4	4,076	437	6.5
BHAN PUR	0000157	444	815	432	383	9.3	31	8	12		756	59	7.1
GOPI RAI	0000158	291	333	175	158	5.6	11	1	1		203	130	6.3
KASSOKI	0000156	544	746	404	342	19.4	86	16	4		712	34	6.1
KIRANWALI	0000160	161	1,175	597	578	9.6	52	11	12	1	1,175		7.0
KOT BHUTTA	0000161	809	627	330	297	17.1	50	8	16		503	124	5.9
PURAN PUR	0000159	634	817	414	403	12.6	43	12	14	3	727	90	6.3
*KOT PURIAN PC			5,987	3,203	2,784	12.0	303	118	85	4	5,653	334	6.7
ALLA DITTA	0000239	1035	1,218	616	602	11.7	58	20	21	1	1,125	93	6.3
BHARAIAN	0000240	736	586	326	260	10.1	20	10	11		512	74	6.2
GHAIBI	0000241	743	914	500	414	6.1	25	8	4		893	21	6.7
KOT PURIAN	0000238	826	1,173	625	548	12.9	70	26	10		1,103	70	7.2
LANJ	0000242	1516	2,096	1,136	960	14.6	130	54	39	3	2,020	76	6.8
*MAAN PC			10,184	5,308	4,876	17.1	999	252	220	19	9,650	534	6.8
HARDO MOGHAL CHAK	0000045	1817	3,460	1,810	1,650	18.0	205	113	95	5	3,262	198	6.8
KOT SHERA	0000046	1226	3,541	1,844	1,697	13.2	223	42	45	3	3,352	189	6.9
MAAN	0000286	1963	3,183	1,654	1,529	20.4	271	97	80	11	3,036	147	6.7
*MANDIALA MIR SHIKARAN PC			10,632	5,719	4,913	17.1	746	223	208	33	9,661	971	6.8
FATEH WALA	0000082	318	995	513	482	21.3	83	33	25	3	826	169	6.8
MANDIALA MIR SHIKARAN	0000079	775	1,591	860	731	15.2	119	27	22		1,529	62	6.8
MAUSHERA SANSI	0000077	1104	6,525	3,547	2,978	18.5	470	139	149	29	5,834	691	7.1
MISRI MAINI	0000081	436	723	393	330	13.6	48	14	7	1	675	48	5.8
NAROKE	0000080	851	798	406	392	7.7	26	10	5		797	1	5.7
*MELO WALA PC			13,869	7,188	6,681	28.1	1,442	539	537	84	13,246	623	7.0
CHIAN WALI	0000125	685	2,953	1,550	1,403	27.7	388	66	63	3	2,897	56	7.0
DHILLAN WALI	0000124	1083	5,337	2,799	2,538	29.6	528	212	265	48	5,035	302	6.9
KHAN PIARA	0000123	520	1,083	553	530	21.0	85	44	28		1,065	18	6.9
MELO WALA	0000120	1067	3,972	2,005	1,967	29.5	411	196	165	33	3,829	143	7.2
THATTA CHILWA	0000122	280	195	93	102	19.5	12	10	4		134	61	6.1
THATTA PANJ HATTAN	0000121	211	329	188	141	17.6	18	11	12		286	43	7.0

Panel B: 2020 Mouza Census Questionnaire



Government of Pakistan
Pakistan Bureau of Statistics
(Agricultural Census Wing)

Form-11
Mouza Census 2020

Identification & Status of Mouza — **Part-I**

01- District :		02- Tehsil :	
03- Kanongo Circle :		04-Patwar Circle:	
05- Mouza Name : S.No.			
06- Hudbast No.		07- Permanent Code :	PCO / ACO
08- Union Council/ Ward Name & No.:			

09- Status of Mouza : [1] Rural [2] Urban [3] Partly Urban [4] Forest [5] Un-Populated — 10- Total Number of Settlements in Area of the Mouza

11- Area of Mouza (Acres) for 2018-19 — 1- Total: _____ 2- Cultivated: _____ 3- Uncultivated: _____ 4- Total Area Under Settlements _____

Important Crops & Irrigation Sources — **Part-II**

01- Major Crops of Mouza	[1] Wheat	[2] Rice	[3] Cotton	[4] Sugarcane	[5] Maize	[6] Pulses	[7] Orchard	[8] Vegetables
02- Sources of Irrigation	[1] Canal	[2] River	[3] Tubwell/ Well	[4] Ravine/ Stream	[5] Spring/ Karez	[6] Arid	[7] Flooding Torrent	[8] Others

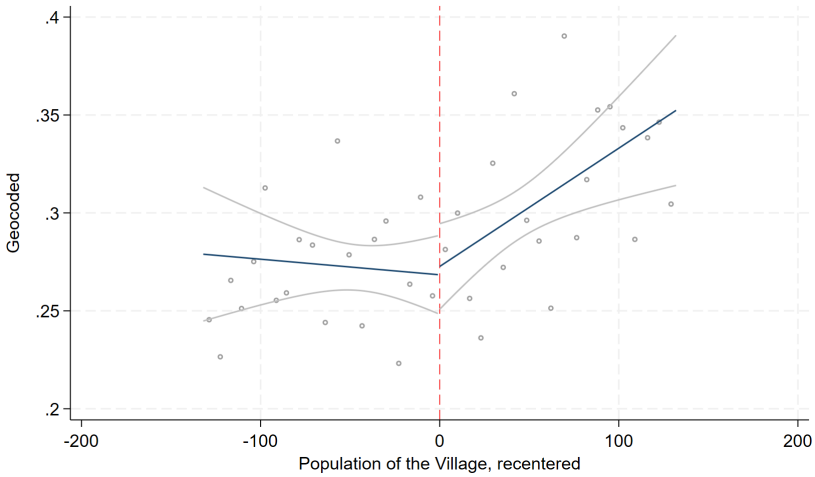
03- Water Course Improvement Scheme — Canal Irrigation [1] Yes [2] No — In Case of "Yes" 1- Total Number of Water Courses ___ 2- Number of Improved Water Courses ___ 3- Total Length of Improved Water Courses (Meters) ___

04- Use of Modren Tech. for Irrigation — [1] Drip System No. ___ [2] Sprinkler System No. ___ [3] Others (Specify) No. ___ [4] None

05- Depth of Water Table in Mouza ___ Feet

Note: The figure presents two sample data sources used in the analysis. Panel A reports the page of 1981 census list of urban localities of Gujranwala district, while Panel B presents the page of questionnaire of 2020 Mouza census.

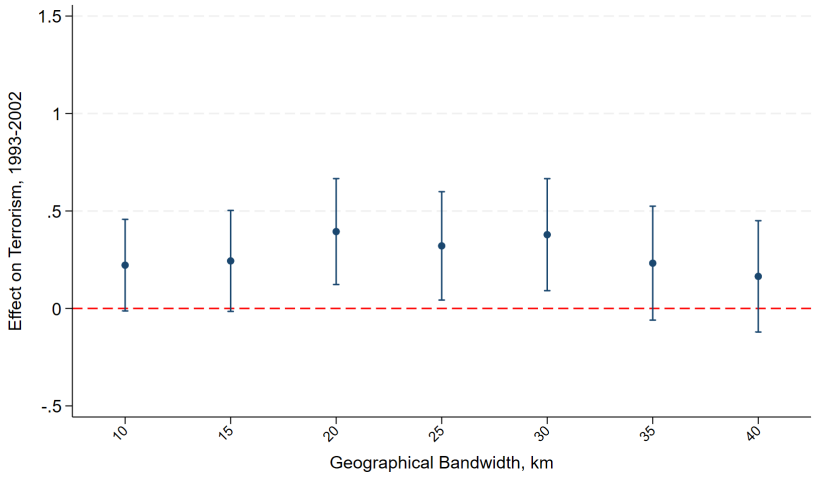
Figure B2: Sample Selection due to Coordinates Availability



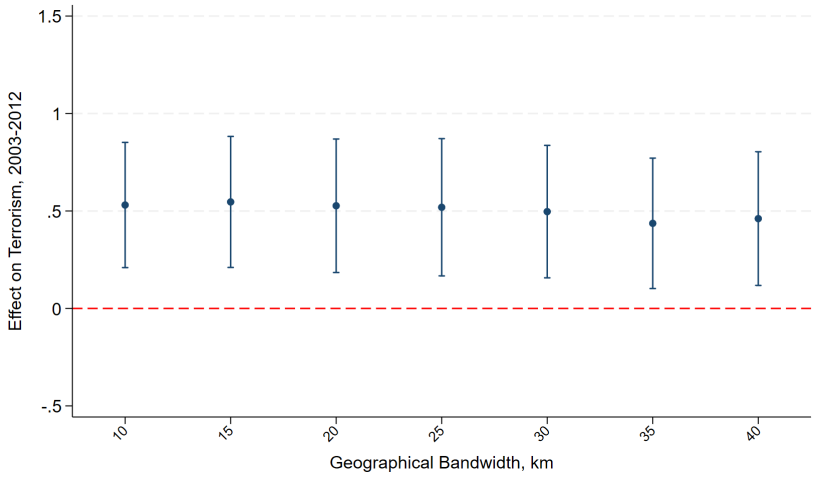
Note: The figure depicts regression discontinuity plots for the analysis of the sample attrition due to coordinates available. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. The dependent variable is an indicator of the village being geocoded using our algorithm. MSE-optimal bandwidth is selected.

Figure B3: Robustness for Choice of Geographic Matching Bandwidth

Panel A: 1993-02

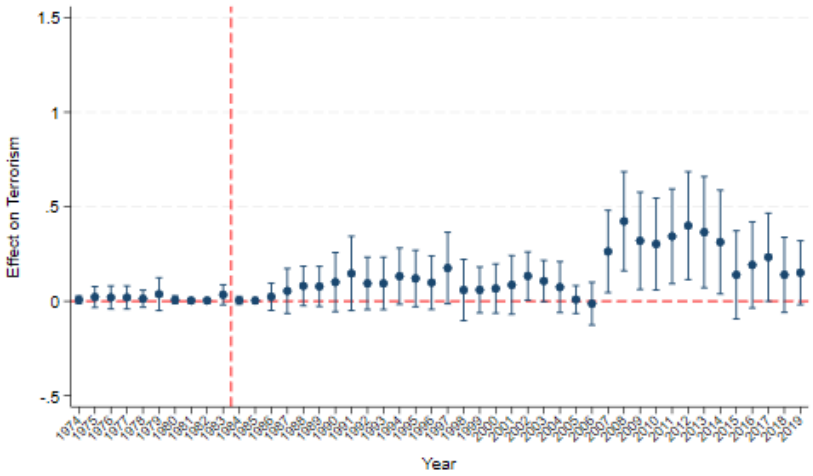


Panel B: 2003-12



Note: The figure presents the robustness of the main results to the different choice of geographical bandwidth. The baseline regression with log number of terrorist attacks in 1993-02 and 2003-12 are estimated with terrorist attacks merged to the villages using the geographical bandwidth of X km.Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator and 90% CI are reported. Regressions are estimated on MSE optimal bandwidth and control for a local polynomial of degree 1 on the both sides of the threshold.

Figure B4: Robustness to the Choice of Time Period



Note: The figure depicts regression discontinuity plots for the analysis of the effect of school construction on terrorism in each year since 1974. The running variable is the population of the village per 1981 census that if exceeds 300 people would result in school construction. The dependent variable is the log number of terrorist attacks in the year. Red line marks the introduction of the program. MSE-optimal bandwidth is selected. Robust standard errors based on heteroskedasticity-robust nearest neighbor variance estimator and 90% CI are reported. Each specification controls for the local polynomial of degree 1 on the both sides of the threshold.

11. Appendix C. Prompts and LLM Analysis.

Exact prompt for table digitizing with Mistral AI.

```
You are extracting data from a page of the 1981 Pakistan Population Census.
You will receive the OCR output (in Markdown) of one census page. The page contains a table
of localities (mauza/deh/village) with demographic columns.


Extract ALL data rows. Follow these rules:


1. NAME: Copy the village/locality name EXACTLY as it appears. Preserve spelling and capitalization.
   - Rows starting with * are Police Circle (PC) or Tehsil aggregates - set is_aggregate=true, strip *.
   - Regular village rows: is_aggregate=false.


2. HADBASTI_NUMBER: 7-digit code like "0000134". Null for aggregate rows.


3. AREA_ACRES: integer. Null for aggregates.


4. pop_both_sexes, pop_male, pop_female: integers (remove commas: 1,322 → 1322).


5. literacy_rate: decimal (e.g. 1.9). Null if blank.


6. primary_ry: integer (Primary educated count). Null if blank.
```

```
7. middle: integer. Null if blank.


8. matric_above_male, matric_above_female: integers. Null if blank.


9. muslim: integer population count. Null if blank.


10. others: integer (non-Muslim). Null if blank.


11. household_size: decimal (e.g. 5.3). Null if blank.


For any blank or missing cell, use null. Do not infer or interpolate.
If a character is ambiguous in a numeric column, use context (census data) to pick the most plausible
↪  digit.


Return ONLY valid JSON - no markdown fences, no explanation:

    "rows": [
    "name": "THOLAGH",
    "hadbasti_number": "0000134",
    "area_acres": 1231,
    "pop_both_sexes": 96,
    "pop_male": 49,
    "pop_female": 47,
    "literacy_rate": null,
    "primary_ry": null,
    "middle": null,
    "matric_above_male": null,
    "matric_above_female": null,
    "muslim": 96,
    "others": null,
    "household_size": 5.3,
    "is_aggregate": false

    ]
```

Exact prompt for textbook analysis.

```
PAGE_SCHEMA_HINT = """Return ONLY a JSON object, no prose, no markdown fences, with exactly these keys:

    "language": "english" | "urdu" | "mixed" | "other" | "blank",
    "english_text": "<faithful English rendering of ALL readable text on the page. If already English,
    transcribe it. If Urdu, translate it. Preserve meaning, not layout. Empty string if blank.>",
    "page_kind": "cover" | "title" | "toc" | "content" | "exercise" | "blank" | "image_only",
    "year": "<4-digit publication/edition year if visible on THIS page, else empty>",
    "subject": "<school subject if clearly stated on THIS page, e.g. 'Islamiat','English','Urdu','Social
    Studies','Mathematics','Science','History', else empty>",
    "mentions_islam": 0 | 1,
    "islam_intensity": <integer count of distinct Islam-related references on the page>,
    "islam_category": "theological" | "ritual" | "historical" | "civic_national" | "incidental" |
    "none",
    "islam_evidence": ["<short verbatim English excerpts that justify the flag>"],
    "mentions_violence": 0 | 1,
    "violence_intensity": <integer count of distinct violence references>,
    "violence_category": "war_battle" | "physical_harm" | "punishment" | "conflict_struggle" |
    "metaphorical" | "none",
    "violence_evidence": ["<short verbatim English excerpts that justify the flag>"]

Coding rules:
- "mention of Islam" = explicit reference to Islam, Muslims, the Prophet, Quran, Islamic
↪  practice/figures/history. Generic mention of God in a non-Islamic frame is NOT islam unless context
↪  is Islamic.
- "mention of violence" = depiction or reference to physical force intended to hurt, kill, damage: war,
↪  battle, killing, beating, punishment, armed conflict, martyrdom involving violence. Sports/figurative
↪  "struggle" = metaphorical, set intensity low and category accordingly but you MAY still flag if
↪  clearly violent imagery.
- Be conservative and literal. Base flags only on what is actually on the page. Evidence excerpts must be
↪  real text from the page (translated to English), max ~15 words each, max 4 items.
"""

SYSTEM = ("You are a meticulous research coder digitizing and content-coding pages of "
          "school textbooks for an academic dataset. You transcribe/translate faithfully "
          "and code conservatively. You output strict JSON only.")
```