Childhood Exposure to School Expansion and Changes in Inter-Caste Marriage in India

Nitin Kumar Bharti*, Luca Maria Pesando†,
Koyel Sarkar‡, Neha Bailwal §

Abstract

Marriage in India is largely controlled by hanging norms such as arranged marriage and sociocultural markers such as the caste system. Caste endogamy, i.e., marrying within the same caste, remains one of the strongest pillars of the caste system in the Indian society, with close to 86% of endogamous marriages in the 2020 marriage cohort. This study explores whether childhood exposure to dynamics of school expansion in rural India had any causal impact on the increase in ICM exploiting variation in school openings across different locations at different times. To do so, we rely on georeferenced information from three large-scale datasets, the District Information System for Education (DISE), the Indian Census 2011, and the National Family Health Survey (NFHS) 2014-15 and 2019-21. A one standard deviation (SD) change in school openings (per village) increases ICM by 5.67%. Exploring the underlying mechanisms, we do not find completed years of education to be the driver, suggesting contact theory, delayed age at marriage, and/or education assortativity are more plausible mechanisms. These results indicate that education, development, and broader modernization forces can be a driver of family change in India, albeit their relevance may be lower than in other low- and middle-income countries (LMICs).

Keywords: Intermarriage; caste; schooling; development; modernization; India.

*Email: nkb2039@nyu.edu. New York University - Abu Dhabi.
†Email: lucamaria.pesando@nyu.edu. New York University - Abu Dhabi. **Corresponding author.**
‡Email: koyel.sarkar@nyu.edu. New York University - Abu Dhabi.
§Email: neha.bailwal95@gmail.com. Indian Statistical Institute Delhi.

1 Introduction

Divisions along caste lines remain a strong defining feature of the Indian society. While several traditional systems of social stratification such as slavery, racism, gender, and class are undergoing relatively rapid changes across the world, the caste system seems to retain its power and to keep regulating the lives of over 1.4 billion people irrespective of religion and ethnicity (Gundemeda, 2020; Vaid, 2014). One of the key characteristics of the caste system is the closed system of marriage or "caste endogamy" (Davis, 1941). Inter-caste marriages (ICM) are still relatively rare in both rural and urban areas in India. According to Caldwell et al. (1998, p. 146), even though "some erosion of arranged marriage has begun [...] and an increase has occurred in cross-caste marriage," these marriages still tend to be between castes of a similar hierarchical level (Caldwell et al., 1983). Despite increases in ICM and uneven prevalence across states, the desire to marry within caste holds strong (Banerjee et al., 2013). Open questions remain about the kind of societal changes and/or policies that may contribute to weakening the caste-endogamy system. Goli et al. (2013) suggest that increases in mixed marriages in India are associated with changing socioeconomic and cultural values, alongside the impact of Western education and economic diffusion. Thinking about the powerful role of education in providing knowledge, helping develop abilities to understand new and different points of view, and bringing together people from different socioeconomic backgrounds, one would expect education to be a rather powerful factor in weakening caste boundaries.

Over the past 75 years since Independence, the education sector in India has expanded massively (Joshee, 2008). Successive governments have worked to ensure maximum enrolment and attendance in primary and secondary schools, in an effort to achieve a knowledge-based economy (Mander and Prasad, 2014). Recent data from the Ministry of Education suggest that the number of schools has grown more than ten times since Independence, from about 141,000 to over 1.5 million (Ministry of Education, 2021). To contribute to these already ongoing efforts, in 1994 the District Primary Education Programme (DPEP) was approved as a governmental scheme to expand, universalize, and transform the quality of primary education in India, following a massive push on the part of the international community to attain the Millennium Development Goal (MDG) number 2, i.e., to attain universal basic education. While the policy discourse has then shifted from the MDGs to the Sustainable Development Goals (SDGs) — alongside a progressive focus from quantity to quality of education — the implications of massive schooling expansions in India for dynamics of family formation are yet to be thoroughly assessed.

This study relates the above two pieces by exploring whether childhood exposure to school openings may be a powerful factor underlying the increase in ICM in India. Using georeferenced longitudinal data on school openings from the District Information System for Education (DISE) merged with the Indian Census 2011 and two waves of the National Family Health Survey (NFHS), we address this question leveraging variation in school openings across different locations at different times. We do so by focusing on rural areas of India – those that experienced some of the most massive "positive shocks" in terms of school openings – as well as focusing on men.[1]

We find that a one-unit increase in the number of primary schools within a village in a 5 km NFHS cluster during an individual's primary school-going age (0-9 years) leads to an increase in education and age at marriage for men by 0.6 years and 0.54 years, respectively. Additionally, the effect of primary school expansion on ICM is positive and statistically significant: a one standard deviation (SD) increase in school expansion raises ICM by 5.67 % of the sample mean. Increased years of education and delayed age at marriage do not appear to be potential channels, suggesting that contact theory may provide a more plausible explanation.

Assessing the implications of schooling expansion on inter-caste marriage is fundamental for at least two reasons. First, inter-caste marriage is a domain that has shown inertia over the past century, leading to a reinforcement of the common perception that family forms and structures in India only weakly respond to development, urbanization, and modernization forces that are pervading low- and middle-income countries (LMICs) (**?**). This aspect has key implications for policymaking, as it would suggest that social policies targeted towards families would do little to shape intrinsic features of the family system. Conversely, should schooling have the — at least minor — potential to affect socio-cultural markers of the Indian family system, then policymakers may push towards redirecting resources to promote, strengthen, and sustain human capital accumulation across the life course and across all strata of society, thus identifying a clear and effective policy lever.

Second, it is undeniable that the persistence of the caste system and, especially, the sustained high prevalence of caste endogamy in marriage has huge implications for societal inequalities in that it minimizes social mixing and integration, reinforcing societal homogeneity and contributing to the reproduction of intergenerational inequalities (Blau et al., 1984; Golebiowska, 2007; Goli et al., 2013; Munshi, 2017; Munshi and Rosenzweig, 2006).

[1]Due to the primarily feminine nature of migration in India, focusing on women would threaten the validity of our identification strategy relying on school openings during childhood.

According to Sharma (2019), caste endogamy provides the foundation to reinforce and sustain the edifice of caste-based society in South Asia, affecting especially Hindus rooted in a rigid caste system (Saroja, 1999). In a way similar to racial or ethnic intermarriage (Kalmijn, 1998; Qian and Lichter, 2007; Schermerhorn, 1970), ICM may signal a significant lessening of "social distance" between minority and majority groups (Song, 2009; Sharma, 2019). In turn, higher social integration and societal heterogeneity correlate with greater solidarity and trust, as well as higher social mobility (Roth and Peck, 1951; Schwartz et al., 2016). Therefore, our study has profound implications for the study of intra- and intergenerational social mobility in the Indian context, a topic that is of key relevance in light of India's role in the geopolitical and sociodemographic landscape.[2]

This paper provides a series of important contributions to the relevant literature. First, the extant literature on education and ICM in India mainly focuses on the role of parents' education and its association with ICM (Ray et al., 2020; Sharma, 2019). Our study contributes to this discussion by highlighting how a spouse's exposure to school expansion during childhood (men, in this case) can influence ICM in rural India. Second, while micro-level studies in India and other countries show a positive link between expanding education and changing marriage practices, including ICM (Sarkar, 2022; Dommaraju, 2010; Medhe, 2019; Sharma, 2019; Singh et al., 2023), it remains uncertain whether there is any causal relationship between expanding education and changes in culturally rooted aspects of marriage. Similar strategies leveraging school-supply "shocks" have been adopted to study polygamy in Cameroon (André and Dupraz, 2023), child health in Taiwan (Chou et al., 2010), wages and education in Indonesia (Duflo, 2001) and India (Khanna, 2023), and choice of school track in France (Garrouste and Zaiem, 2020).

Third, we offer an updated picture of prevalence and trends in ICM using some of the latest data and adopting a marriage-cohort approach tracing cohorts back to the 1980s. Lastly, the study also contributes to the growing literature on early childhood intervention and its long-term social and economic implications (Bharadwaj et al., 2013; Boucher et al., 2020; Bharti and Roy, 2023). We enhance this literature by investigating how exposure to school expansion during men's formative years affects marriage market outcomes such as ICM. Our findings underscore the importance of targeted interventions in early childhood to weaken caste boundaries in rural India.

The study is structured as follows. Section 2 provides contextual information on recent

[2] Enough to think that, as of April 2023, India overtook China as the most populous country in the world (UN-DESA 2023).

trends in changes in families in India, as well as on the relationship between individual levels of education and ICM. Section 3 describes data, sample, and variables. Section 4 provides descriptive statistics on the analytical sample. Section 5 outlines the empirical strategy. Section 6 discusses the results, highlighting various layers of heterogeneity. Section 7 explores potential mechanisms, while Section 8 concludes with discussions and policy implications.

2 Background and Context

2.1 Family change in India

Marriage in India is almost universal as 90% of men and women get married by 29-34 years of age and less than 2% remain unmarried by the age of 49 years and above (Bhagat, 2016). Marriage is also a one-time event in the lives of most Indians, as the divorce rate has never gone above 1% over the last censuses (Jones, 2017; Dommaraju, 2016). As such, marriage has remained and continues to be one of the most important pillars of the Indian society. Marriages are typically arranged by the family with careful consideration of caste, religion, language, and other ethnic and socioeconomic attributes of both spouses. Familial involvement is high because of the patriarchal, patrilineal, and patrilocal structure of Indian families, which are usually extended with several sub-units living together. As such, marriage is treated more like an alliance between two families rather than two individuals (Béteille, 1991). Thus, the reproduction of the social structure of the family, in terms of socioeconomic status, caste, honor, reputation, etc., remains crucial. Premarital courtships and love marriages used to be largely discouraged as these often led to"irrational" partnering transcending these well-formulated schemas. Dowry in the North and bride price in the South are standard practices associated with these norms.

Most of these norms continue to operate in Indian marriages today; however, there have been some notable changes over time and space. Mostly due to the rise in education levels, employment, and urban living, family ties have been gradually loosening, along with changes in beliefs, norms, and practices (Ross, 1961; Kannan, 1963; Corwin, 1977; Thornton, 2001). Most changes can be seen in the disintegration of larger joint families into smaller nuclear families (Allendorf, 2013), increased involvement of couples in marriage-related decisions, and the importance given to love and compatibility even within arranged marriage configurations (Caldwell et al., 1983; Uberoi, 2006; Fuller and Narasimhan, 2008). Recent findings evidence a shift from "fully" arranged marriages to "jointly" arranged marriages, a decline in consanguineous marriages, a modest increase in ICM, and spouses meeting before the wedding day, making India a hybridized model of both modern and traditional practices (Allendorf and Pandian, 2016). Furthermore, recent research also highlights a revival of intergenerational co-residence, signaling the persistence of strong family ties (Breton, 2021; Esteve and Reher, 2021). This is relevant in that living with parents and older generations could hinder shifts in couple formation from traditional norms to modern configurations (Ray et al., 2020; Goli et al., 2013).

These changes have been more rapid in some regions than others. For instance, self-arranged or love marriages are more common in the East (Donner, 2002; Allendorf, 2013; Sarkar and Rizzi, 2020), ICM and love marriages are more prevalent in the North-east (Sarkar and Rizzi, 2020), while meeting potential spouses over online dating apps (Chakraborty, 2012; Tritzmann, 2011) and experiencing jointly arranged marriages are more common in the South (Fuller and Narasimhan, 2008, 2013; Netting, 2010).[3] These changes suggest that as couples start to break free from the conventional marital arrangements where parents are the main decision makers, conventional norms may also weaken, caste endogamy being one of them. However, although there has been a modest increase in ICM, mostly associated with self-arranged or love marriages (Allendorf and Pandian, 2016), caste boundaries remain binding, as parents and family approval continue to be one of the key driving factors behind these arrangements (Sarkar and Rizzi, 2020).

Although marital arrangements are shifting slowly, a major transformation has occurred in age at marriage over the last decades, which has been identified by many scholars as a key driving factor behind marriage change in India (Sarkar and Rizzi, 2020; Caldwell et al., 1983; Prakash and Singh, 2014). Delayed marriages not only provide more time to choose one's spouse independently, but also suggest less resistance from families in cases of unconventional matchings, such as ICM (Ghimire et al., 2006). The mean age at marriage was 18 years for women and 23 years for men in the 1991 census (Bhagat, 2016), an estimate which increased by two years in the 2011 census, raising the median age at marriage to 20 years for women and 25 years for men. Numbers vary greatly across states and regions: marriage ages are much higher in the Southern states, Goa, Kerala, Tamil Nadu, and Karnataka compared to the Northern states due to differences in female autonomy and demographic behavior (Jones, 2017; Bhagat, 2016; Dyson and Moore, 1983; Karve, 1968; Trautmann, 1981).

[3] Some ethnographic evidence of less normative marriage trends has also been found in Gujarat (Netting, 2010), Andhra Pradesh (Still, 2011), Ladakh (Aengst, 2014), Haryana (Chowdhry, 2004), and Delhi (Mody, 2008).

2.2 Prevalence and trends in inter-caste marriage

Encouragement of ICM has been one of the tools of Indian governance systems to usher in multiculturalism, eradicate untouchability, and facilitate social integration. Inter-caste marriages were written into law in India by the Special Marriage Act of 1872 and recognized by the Hindu Marriage Act of 1955. In 2011, the Supreme Court of India declared that ICMs "are in the national interest" and provide a unifying factor for the nation, as there has never been a bar on inter-caste or inter-religious marriages in independent India.[4] These statements even translated into policy initiatives, such as the *Dr. Ambedkar Scheme for Social Integration through Inter-Caste Marriages 2013*, with a monetary incentive given to inter-caste couples. Regardless of such policies, efforts, statements, and public acknowledgments, most marriages continue to take place within the same caste, due to lack of effective implementation and widespread opposition in society.[5]

While the vast majority of marriages are endogamous and changes are slow, Figure 1 shows a close-to-monotonic increase in ICM from 10.4 in the earliest marriage cohort (1981-85) to 14.8 in the latest marriage cohort (2016-2020) using two waves of NFHS data from the whole of India. A 4.4 percentage-point increase over 40 years may be deemed negligible, yet it corresponds to a 42% increase.

Goli et al. (2013) provide the most updated and comprehensive overview of prevalence of ICM in the Indian society through the 2005 Indian Human Development Survey (IHDS). They found higher prevalence of ICM in states such as Punjab (12%), West Bengal (9%) and Gujarat (8%). They observed, overall, a higher likelihood of mixed marriages in urban areas and among the economically, educationally, and culturally advanced, and recognized that the prevalence of mixed marriages may increase with changing socioeconomic and cultural values, alongside the impact of Western education.[6] Focusing on the second round of the IHDS, Sharma (2019) found the proportion of women marrying beyond caste boundaries to be higher for younger (less than 19) as well as older (above 30) brides. Lastly, thinking about factors underlying the persistence of "casteism," Saroja (1999) highlighted prejudices and myths against mixed marriage, as well as adverse social forces such as increasing competition for jobs, as factors contributing to reinforcing caste endogamy.

[4]In 2018, the Court reiterated that marriage between consenting adults is fully legal, notwithstanding their caste, and underlined the rights of adults to choose their partners as well as the need for society to learn to accept inter-caste and inter-faith marriages.

[5]For instance, 41% of adults in Delhi and 62% of adults in UP were in favor of laws banning intermarriage between high and low caste (Hathi, 2019).

[6]Munshi and Rosenzweig (2009), through their primary data, also find similar estimates: among 25-40-year-olds, ICM was 7.6% in Mumbai in 2001, 6.2% in South Indian tea plantations in 2003, and 5.8% for the rural Indian population in 16 major states of India in 1999.

2.3 Education and ICM

A selected number of studies have explored the relationship between own or household level of education and likelihood of ICM. Using the IHDS 2011-12, Sharma (2019) found that exposure to education and, specifically, parental (mother of the bride) education, is associated with more freedom in spousal choice and, in turn, a higher likelihood of ICM. Conversely, she found no significant associations between the bride's level of education and her choice of mate once other factors are controlled for. Nonetheless, higher education of the bride is associated with a shift away from arranged marriage and the adoption of a so-called "middle path," whereby brides create a balance between self-chosen and parentally arranged marriage while simultaneously retaining their loyalty towards their society, community, religion, and parents, in line with Allendorf and Pandian (2016). Addressing a similar question, Ray et al. (2020) found that education levels of spouses themselves do not bear any association with the likelihood of their own marriage being an ICM, yet the level of education of the husband's mother has a positive and statistically significant association with the likelihood of ICM, thus underscoring the still prominent relevance of the institution of arranged marriage as well as parental approval. The authors found that one SD increase in years of education of the husband's mother is associated with a 10% increase in the probability of ICM over the sample mean. This said, causal analyses on the role of education —- be it own education or parental education — on family-related outcomes in India remain scarce.

Education has the potential to mitigate rooted prejudices and stereotypes, and educational institutions can serve as platforms for social mixing and social integration, in turn shaping perceptions and attitudes towards the "other" (Ray et al., 2020). Scholars have formalized these ideas outlining four theoretical perspectives that may be applicable to our context. First is the so-called "enlightenment perspective," which suggests that the roots of prejudice and intergroup antagonism are uninformed worldviews, thus less education and lower cognitive ability are linked with starker prejudices, attitudes, and stereotypes (Wodtke, 2012). Scholars have found that more education gives people the ability to better analyze and understand different points of view, processing information through value systems that go beyond their own (Deary et al., 2008; Schoon et al., 2010), thus conducing to more tolerance towards the "other" by imparting knowledge about minority groups, teaching how to recognize prejudice, and outlining history's injustice towards minority groups (Hathi, 2019; Hodson and Busseri, 2012).

Second is the "cultural adaptability hypothesis," which is closely tied with the enlightenment perspective yet usually applied to inter-language exchange. This postulates that individuals with higher levels of education and cognitive ability tend to have stronger motivation to adapt to foreign cultures and are more inclined to change the behavior patterns used in their original culture to adapt to the circumstances of the new environment, while lower-educated individuals appear more "passive" when adapting to foreign cultures (chun Lin et al., 2012; Ray et al., 2020).

Third is "contact theory," popular in sociology and psychology, suggesting that intergroup contact may effectively reduce prejudice and conflicts between majority and minority groups, leading to more acceptance and social mixing (Allport, 1979). As such, institutions such as schools would be conducive to social mixing by boosting intergroup contact.

Fourth, the "assortative mating theory," which is less tied to the idea of social mingling and more related to "objective" spousal marriage-market characteristics, postulates that in a group with average education level above the average education level of the corresponding population, a more educated individual may "marry in" and education may be negatively associated with intermarriage for that group. The outcome can go in either direction and one may observe a positive, negative, or no relationship between education and intermarriage depending on a particular group's characteristics, as outlined by Ray et al. (2020) and Furtado (2012).

3 Data

To assess the impact of school expansion on inter-caste marriages in rural India, we employ three separate datasets: (1) National Family Health Survey (NFHS) for the years 2015-16 and 2019-21, (2) District Information System for Education (DISE) data for 2011, and (3) Census data from the Office of the Registrar General and Census Commissioner, Government of India, for the year 2011.

3.1 Data on inter-caste marriage

The NFHS is a nationally representative household survey offering comprehensive information on various aspects such as population, health, marriage, fertility, and nutrition. The household selection follows a stratified two-stage sampling approach. In the initial stage, Primary Sampling Units (PSUs), also referred to as clusters (equivalent to villages) are selected from the 2011 Census list using probability proportional to size. In the subsequent stage, an equal number of households (20) are randomly chosen within each PSU.

We use two rounds, namely NFHS-4 conducted in 2015-16 and NFHS-5 conducted in 2019-21, to measure inter-caste marriage among couples for our analysis. The survey interviews all women aged 15-49 and men aged 15-54 within the selected households. Our analysis focuses exclusively on men, considering that women's village of residence tends to change following marriage.[7] We use the couple-level file in the NFHS, asking both husbands and wives the same questions about marriage. Our dependent variable is whether the spouses belong to different caste groups. We restrict the sample to rural areas, with men above 22 years and older, to ensure the majority of them have attained their highest level of education.

The NFHS also provides the geographical coordinates of each cluster. However, these coordinates are randomly displaced within a 5-kilometer radius to ensure respondent privacy.[8] Since we do not know the exact cluster, we create a 5km radius buffer zone around the given geographical coordinate of a cluster and create the main explanatory variable at this buffer zone level.[9]

3.2 Data on schools

We rely on the DISE-2011 data to look at the expansion of schools over time. DISE is an annual data set that covers a universe of schools (nearly 1.2 million) all over India, with school-level information, such as the year of opening of schools, type/levels of schools (primary, secondary, higher secondary), school management structures (Department of Education, Tribal/ Social Welfare, local body, private aided, etc.), and school's highest and lowest grade levels. For our analysis, we focus on primary schools only.

Furthermore, DISE incorporates geographical details about the school, including district, block, and village names, which link this data with the 2011 Population Census. However, without unique village-level codes in the DISE dataset, the linking process involves

[7] Evidence from Indian Human Development Survey 2011-2012 (IHDS II) data suggests that migration among women is higher in rural India with more than 80 percent of women reporting that their childhood residence is different from their current village of residence.

[8] The geographic coordinates of 99% of the clusters in the rural stratum were displaced up to 5 km, with 1% being displaced up to 10km. The displacement was restricted to keep the clusters in the same districts.

[9] There is 99% chance that the actual surveyed village is within this buffer zone. Further, to avoid the measurement error, we only keep those villages, with at least 10% of their area falling into the 5km buffer zone.

name-based fuzzy matching. The fuzzy matching of village names is done within a given district and subdistrict, to be precise. Further, a thorough manual examination is carried out to keep precise matches.

This merged DISE-Census data is further combined with NFHS 5km zone data using unique village-level census codes, and we keep only the relevant villages, i.e., falling within the 5km buffer zone. This provides us with the school-level information within the buffer zones. Using the year of opening of schools, we create an annual balanced panel at the buffer zone ("DHS-5km cluster") of the number of primary schools ranging from 1950 to 2012.

Lastly, we combine the couple dataset with the above-created panel dataset to compute our main regressor — change in the number of schools in the husband's first 9 years of age.[10]

3.3 Village level characteristics

We obtained village-level characteristics from the latest round of Census data, which provides comprehensive information on population composition and village amenities. Our village-level controls are constructed using primary census abstract and village directory data. To conduct our analysis, we merged the Census data with DISE data at the village level. It is important to note that there is no unique village identifier available that can be used to merge Census and DISE files. Therefore, we use fuzzy matching techniques to merge 2011 Census data with DISE data based on the state, district, block, and village name. Total number of Census villages was 6,40,950 in 2011. However, 43,330 villages are not populated. We drop all such villages from our analysis. Since DISE covers only the villages with at least one school, all the Census villages without the primary school also get dropped from our sample. Finally, based on fuzzy matching, we could match 404,926 villages. Subsequently, we merge this school and village-level data with NFHS by utilizing the geographical coordinates.

4 Descriptive Statistics

In the sample, prevalence of ICM is far lower than marriage endogamy. Overall, the prevalence of ICM exhibits an upward trajectory between 1981 and 2021, driven by all caste cat-

[10] It is simply the difference between the number of schools at age 9 and the number of schools at age 0 (i.e., birth year)

egories, except for a slight declining trend in the General Caste for the last marriage cohort (Figure 1).

Appendix Figure A.I shows cross-tabulations by husband and wife's caste, providing a glimpse into ICMs. As the gap between caste groups increases, prevalence of ICM declines. For example, 83% SC men marry SC women compared to a meager 8.8% marrying to OBC, 5.8% to ST, and only 2% to General caste. Conversely, when the husband is from the SC caste, the inter-caste marriage share shows a declining trend. Similarly, 88% ST men marry within their own caste, 7% to SC, 4% to OBC, and 1.2% to General.

In figure A.II, we descriptively explore the relationship between inter-caste marriage and school exposure across different birth cohorts in rural India. When schools are fewer (i.e., below the median), it indicates lower school exposure. In this case, we observe a consistently lower level of ICM. Notably, the increase in ICM appears more pronounced within specific birth cohorts, highlighting that the impact of school exposure on inter-caste marriages varies across different generations. To sum up, there is a positive association between school exposure and inter-caste marriage. Empirical analyses that follow attempt to identify this pattern causally.

5 Empirical Strategy

To identify the effect of school expansion on inter-caste marriage, we need exogenous variation in school expansion such that individuals are assigned randomly to the change in number of primary schools in a cluster. We use a combination of two exogenous variations to estimate the causal effect. We first exploit variation in school expansion across different birth cohorts within the same cluster. Our second source of variation comes from men in the same birth cohort exposed to different school expansions in different clusters.

5.1 Baseline Specification

Our baseline specification is as follows:

$$y_{idb} = \alpha + \beta post_{db} + \beta_1 post_{d,b-1} + \delta_d + \rho_b + X_i + \epsilon_{idb} \tag{1}$$

where, y_{idb} is 1 for inter-caste marriage for individual i born (and living) in NFHS-5km cluster d, belonging to birth cohort b. Our main coefficient of interest is β, which captures the causal impact of primary school expansion on inter-caste marriage. The covariate $post_{db}$ is a continuous variable capturing the intensity of the primary school expansion for individuals i aged 0-9 years. It is measured as the change in the number of primary schools per village[11] for individual i in the first nine years of the birth (i.e., during age 0 to 9 years). The assumption is that opening new primary schools will impact the cohort that was 0-9 years old (the age of attending primary school); the later cohort will miss the opportunity to attend newly opened primary schools. δ_d and ρ_b are the cluster and birth cohort fixed effects. The covariate $post_{d,b-1}$ is the number of primary schools per village one year before the birth- capturing the impact of the prevalence of existing primary schools on individual i. X_i includes individual-level controls like dummy for religion (Hindu and Muslim) and household current wealth.

Our main identifying assumption is that change in the number of primary schools in a cluster is exogenous conditional on cluster and birth year fixed effect. We control for unobserved differences in birth year by including birth year fixed effect (ρ_t). Clusters with large school expansion may be different from clusters that experienced less change in the number of schools. We take into account such unobserved differences in clusters by controlling for cluster fixed effects captured by δ_d.

While we believe that change in the number of primary schools in a cluster is exogenous, there could be a concern that individuals may migrate to clusters with large school expansion. Given the low level of migration for men in rural India, we believe that migration is less of a concern in our setting (Munshi and Rosenzweig, 2016).[12]

6 Main Impact of the Expansion of Primary Schools

This section of the paper presents our main results. Table 1 displays the impact of school expansion on variables of interest — years of education, age of marriage for husbands, age of marriage of wives, and inter-caste marriage — using our baseline specification.

Column (1) of Table 1 shows that a one-unit increase in primary schools in a village in a 5 km NFHS cluster during an individual's primary school-going age (0-9 years) results in an increase of 0.6 years in education on men (8% increase over the mean) at 1% significance level. It is important to highlight that it is not the total impact of the existence of

[11] We normalize the change in the number of schools with the number of villages falling in the NFHS cluster as per census 2011. It is to take into account different sizes of NFHS clusters. It is synonymous with normalizing with respect to an area of the cluster. The other possible contender, being population, has the problem of reverse causality (**cite**)

[12] Data from IHDS reveals that over 90 percent of the sampled households have been residing in their current village of residence for more than 50 years."

schools, but the opening up of new schools conditional on the existing stock of schools. The β_1 coefficient captures the impact of the existing stock of schools per village, which is 0.439 at 5% significance level (6% over the sample mean).

Column (2) of Table 1 shows that a one-unit increase in primary schools in a village in a 5 km NFHS cluster during an individual's primary school-going age (0-9 years) results in an increase of 0.54 years in the age of marriage on men (2.4% increase over the mean) at 1% significance level. However, we do not observe any statistically significant impact on wives' marriage age (Col (4), same table). The subdued effect (compared to education) on the variable related to marriage is not surprising due to families being major decision-makers and strong cultural norms in the rural setup.

We test wives' years of education in Column (3) of Table 1. The coefficient is positive at 0.352 (7% increase over the mean) at 1% significance level, suggesting increasing education-level assortativity.

Column (5) of Table 1 provides the impact of school expansion on inter-caste marriage. Our main coefficient of interest β shows the effect of a one-unit increase in the change in primary schools on the marriage outcome of men exposed to school expansion during the age of 0-9 years, compared to men who did not experience any change in the number of primary schools. The effect of school expansion on inter-caste marriage is positive and statistically significant at 5% significance level (.028). One SD increase in the school expansion increases ICM by 5.67% of the sample mean.[13]

[13]The standard deviation of our main regressor is 0.26; the impact in the ICM for 1SD change is going to be 0.028*0.26=.00737; which is 5.67% of the mean (.00737/.13)

7 Threats to identification

In this section of the paper, we explain strategies implemented to test our key identifying assumptions.

7.1 Time-varying covariates correlated with school expansion

One primary concern with the fixed effects identification approach is that time-varying covariates could be correlated with primary school expansion exposure. In Appendix Figures A.III-A.VII, we present a comparison of pre-trends in variables such as total population, share of SC population, full-time workers share, population share engaged in agriculture, and population engaged in service sector. We find that exposed and non-exposed populations experienced similar trends along these variables for at least five years before and after birth.

7.1.1 Expansion of schools correlated with other infrastructure expansion

The expansion of a particular infrastructure in a cluster could come with the provision of other types of infrastructure. In our sample, primary school expansion exposure correlates with other infrastructure expansion (Appendix Table B.II), such as health, road, and communication facilities. Such simultaneous infrastructure expansion could impact ICM independently (or jointly) with school expansion. It implies our main specification may be capturing the effects of other infrastructure expansion or a combined effect of school and other infrastructure. Restricting the exposure by primary school-going age reduces such concern - as there is no reason for roads and communication to have an impact only during the primary school-going age. However, one could still worry about such bias.

We perform two empirical strategies to remove this concern. First, we conduct a placebo analysis to examine whether changes in other infrastructural facilities impact ICM. We substitute our main explanatory variable with exposure to other infrastructure facilities during the ages 0 to 9. All coefficients are small and statistically insignificant at the conventional level (Appendix Table B.IV). Additionally, we add changes in other infrastructural facilities as control variables and find that the impact of school expansion on ICM remains stable (Appendix Table B.III) as before and statistically significant at the conventional level.

Lastly, we do not find evidence of simultaneous changes in demographic and development characteristics (share of the total population, the share of SC/ST population, sex ratio, and the share of agriculture, among others) with the change in primary schools in a cluster (Refer Appendix Table B.V).

Further, in our extended specification, we add census-level controls, including demographic (population, share of SC, share of ST, share of female), educational (status of other educational institutions), infrastructure (health, road, communication) and economic (share of agriculture, share of manufacturing and share of service) five years before birth, during exposure (0-9) and five years before the marriage year in our main specification and find the magnitude of our coefficient estimates remaining unchanged and statistically significant at 95% confidence interval (Appendix B.VI).

7.2 Selective migration

The migration of households towards clusters experiencing high school expansion would violate the SUTVA (Rubin, 1980), a critical identifying assumption. To avoid this concern, all our analysis is restricted to the population which has always lived in the same place.[14] We also show that migrating men in the survey do not differ from non-migrating along dimensions of religion, caste, and wealth, thus suggesting that attrition is unlikely to be a source of concern (Appendix Table B.VII)

8 Robustness

As a first robustness check, we exclude specific social groups from the sample and examine whether our main coefficient on ICM remains robust. We find that our results remain qualitatively similar when we exclude the ST group from our sample or consider only Hindus who are non-ST, as shown in Column (2) and Column (5) of Appendix Table B.VIII. However, our coefficient of interest loses significance if we exclude Muslims or take Hindus with ST as our sample (Columns (3) and (4) of Appendix Table B.VIII). This suggests interesting heterogeneity across castes, which we explore later.

The next robustness check shows that results are not driven by any possible selection of sample coming from inefficiency in the matching rate. In Appendix Table B.IX, as we move from Column (1) to Column (5), we concentrate on a sample with an improved matching rate. For example, in Column (1), we consider a sample where we could match at least 50 percent of villages in the NFHS cluster with DISE Census data. Similarly, in Column (5),

[14]It is based on the response to the survey questionnaire.

we narrow the sample to cases where we could match 90 percent or more of the villages. We find the coefficient to be quite stable (0.019-0.027) across all columns at 10% significance level, except when we restrict too much in Columns (4) and (5), leading to a drastic reduction in sample size.

We then conduct a series of robustness checks shifting the window of exposure to primary school openings. First, in Table B.XI we report results using change in school openings after age 11, i.e., theoretically not affecting the likelihood of going to primary school anymore (placebo test). Results show, as expected, that change in school openings at later ages do not affect ICM whatsoever, irrespective of the window chosen (12-17, 12-18, 12-19, 11-19). Afterwards, we focus on primary school ages, restricting (0-7 or 0-8) or widening (0-10 or 0-11) the time window of exposure. Our results (B.XII) show that 0-8 and 0-9 (our chosen specification) seem to provide the most consistent results.

9 Heterogeneity

9.1 Heterogeneity by intensity of school expansion

Table 3 examines heterogeneity by intensity of school expansion. We categorize the change in the number of primary schools in a cluster into four categories: less than 25 % change in number of primary schools, between 25-50 % change in number of primary schools, between 50-75 % change in number of primary schools, and more than 75 % change in number of primary schools. We observe that the impact of school expansion on the years of education and age of marriage for husbands and wives is most pronounced and statistically significant in the fourth category. However, we do not find any differential impact on ICM based on the intensity of school expansion.

9.2 Heterogeneity by social groups

We examine the impact of school expansion on ICM for different social groups. In Table 4 and 5, we observe that the impact of school expansion is positive and statistically significant for men belonging to the General and OBC groups. Interestingly, while the husband's years of education no longer show a significant association with ICM, the wife's years of education are negatively associated with ICM among the General castes.[15] While the effect

[15] Once again, the stability of our coefficient of interest, even after including the years of education and age of marriage for both husbands and wives, suggests that these factors are not the potential channels of explanation.

of school expansion for SC is positive (but insignificant), for ST it is negative (insignificant), as shown in Table 6 and Table 7. This could be because ST groups are geographically isolated, and their integration with other communities is much lower compared to other caste groups.

Next, we interact men's caste group with our main regressor to check for any differential impact by caste groups. The coefficients are plotted in Figure 2 at 5% significance level. All the coefficients are significant at 5% level. The magnitude for ST is negative at 0.072, showing that primary school expansion had an overall negative impact on the ICM within ST. The coefficients are positive for all other three caste groups: SC - 0.093, OBC and General - 0.012.

10 Possible Mechanisms

In this section, we explore some possible channels that might be driving our main results on ICM.

10.1 Is the effect driven by increasing years of education?

We start by testing the most obvious channel of gains in years of education (due to the expansion of primary schools). First, we add the husbands' years of education as a control in our baseline specification and find that the main coefficient remains almost the same as before (Column (2) of Table 2; Column (1) shows the baseline results) at 5% significance level,[16] suggesting a lesser role of direct increase of education to be the main channel.

Next, we split the sample by the four completed levels of education of men — no education, primary, secondary, and higher — as captured during NFHS surveys. The coefficient of interest is positive and of similar magnitude (to the baseline specification) — for three (out of the four) groups. The coefficient for those who have studied till the primary level is negative. However, all the coefficients are statistically insignificant even at 10% confidence level (Appendix Table B.X). Once again, the lack of meaningful differential results suggests that gains in education are likely not the main channel.

Finally, we interact men's complete levels of education with our main regressor, to check for any differential impact. Estimated coefficients are plotted in Appendix Figure A.VIII,

[16] Notably, the coefficient on the husbands' years of education is negative at 1% significance level, suggesting that higher education among husbands is associated with lower rates of inter-caste marriage.

with "no education" as the base category. The coefficients on "primary", "secondary", and "higher" are close to zero, showing no differential impact of completed years of education.

10.2 Is the effect driven by the increasing age at marriage?

Increasing age at marriage allows more time to choose one's spouse by oneself and possibly could offer less resistance from families in unconventional matchings, such as ICM (Ghimire and Axinn 2006). Hence, we test this mechanism explicitly.

First, we add the age at marriage of the husband (men) directly in our main specification and observe if there is any change in our main coefficient. Col(4) of Table 2 shows that adding this variable as a control doesn't change our main coefficient's magnitude or significance level. Adding the age at marriage of the wife (Col(5) of Table 2) also doesn't change the result. This exercise suggests that age at marriage may also not be the main channel.

10.3 Is the effect driven by "contact theory"?

Intergroup contact is recognized as one of the most effective means to reduce prejudice and discrimination (Allport, 1979). When members of majority groups engage with individuals from minority groups, it leads to better interaction and understanding, ultimately leading to a reduction in prejudice and discrimination against minority groups.

The literature suggests that childhood, providing the formative years of development, plays a crucial role in shaping individuals' social preferences and choices. For example, increased exposure of children to other ethnicities during their early years is associated with a higher likelihood of forming interethnic friendships (Boucher et al., 2020).

We believe that with the expansion of primary schools, children have had the opportunity to come into contact with peers and teachers from different social groups in their early childhood, potentially leading to long-lasting impacts on their social preferences, behavior, and choices.

Expanding schools as spaces for interaction and contact can have a greater impact in villages where encountering individuals from diverse social backgrounds is otherwise less likely. For instance, in villages characterized by larger total populations and lower diversity, the probability of interacting with someone from a different social background is low. Therefore, schools in such settings can play a crucial role in providing opportunities for individuals to come in contact with someone from diverse social backgrounds.

To test this, we test whether there is any difference in the results between exposure to changes in public and private schools, with the assumption being that public schools may cater to broader socioeconomic strata, thus making it more likely for students to meet across caste lines. As such, we may expect stronger coefficients on ICM within public schools. Results in Table B.XIII suggest that this is not the case and results are similar between public and private schools. Moreover, we check heterogeneity by intensity of exposure to changes in schools, divided into quintiles. Results in Table B.XIV suggest non-linearities in the relationship.

Furthermore, we split our sample in four categories based on total population and village diversity. These categories include low total population and low diversity, low total population and high diversity, high total population and low diversity, and high total population and high diversity. Our findings reveal that school expansion has a statistically significant impact in larger villages with low diversity (Col(1) of Table B.XVII), underscoring the crucial role of schools as spaces for interaction and contact among children from different social groups, especially in settings where contact is otherwise less likely. The same result is observed in Table B.XVI) computing diversity differently, using SC share.

Expansion of schools has higher impact on villages with low diversity as expansion of schools in a such villages will not result in segregation based on caste. Conversely, in villages with a higher diversity, the proliferation of schools may lead to caste segregation.

10.4 Schools as meeting place

10.5 Is the effect driven through parents?

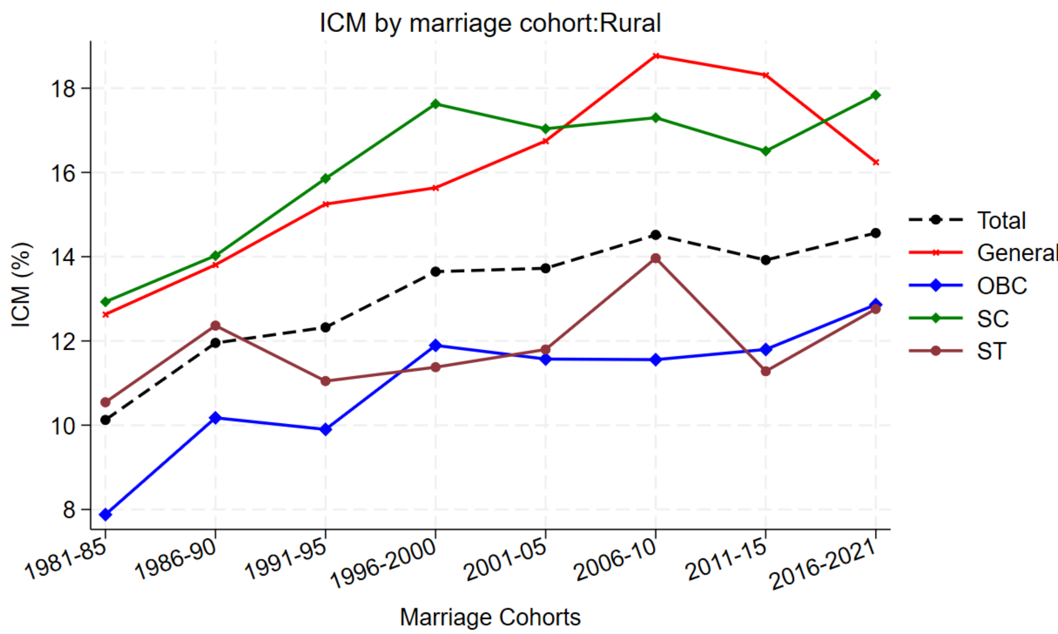
Parents play a crucial role in Indian marriages. One limitation is that we could perform this analysis only when parents are alive and are a part of the same household. We are able to map parents' information (husband father's, and mother's education level) for about 25% of the sample. On this restricted sample, interacting with parents' education level (in 3 categories - illiterate, up to primary level, primary and above), we get no differential impact with fathers' education. [To be completed]

11 Conclusion

We identify the impact of school expansion on inter-caste marriages in rural India. We show that increasing years of education and increasing age at marriage are not the most likely channels. We hypothesize that since contact with other castes (in schools) occurred at an early age, it may have developed a positive attitude towards different caste groups. However, we still have to explore this channel in more detail, which we are planning to present in the next draft.

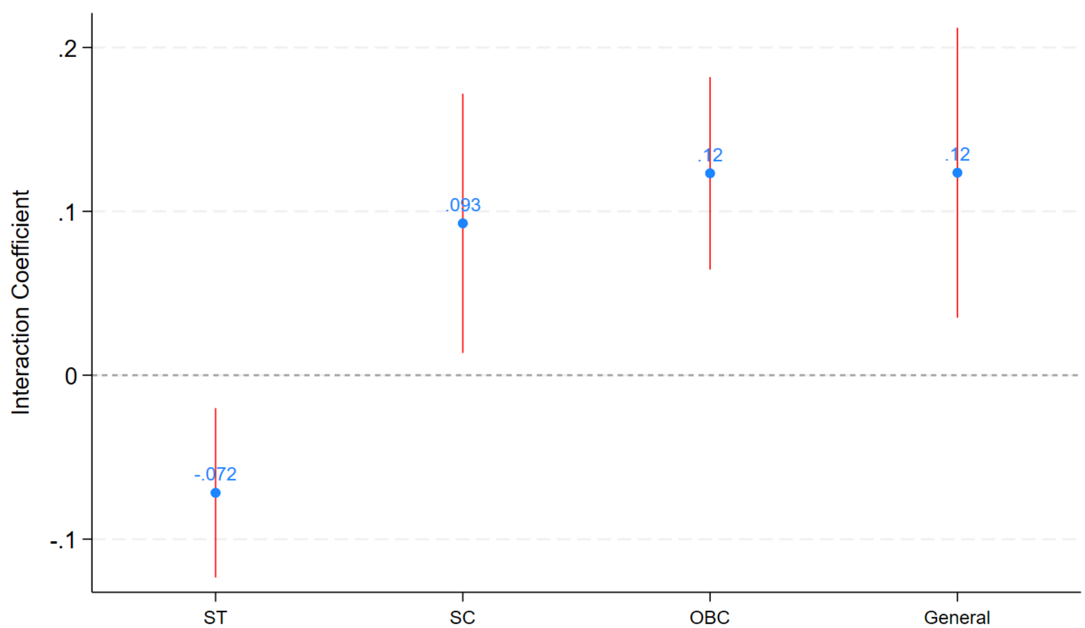
12 Figure

Figure 1: Evolution of ICM



Notes: The figure plots the evolution of inter-caste marriages by birth cohort in rural India from the 1980s till now. It also shows the ICM within each caste category (using the husband's caste). Calculations using NFHS combined sample without using survey design weights.

Figure 2: Differential Impact of Men's Caste Group



Notes: The figure plot coefficient on the dummy of caste group after the ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever - following Equation 1. The dependent variable is a dummy of inter-caste marriage.

13 Table

Table 1: Impact of Schools on Years of Education, Age of Marriage and Inter-Caste Marriage

VARIABLES	(1) h_educ_yr	(2) age_marr_H	(3) w_educ_yr	(4) age_marr_W	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.601***	0.541***	0.352***	0.143	0.028**
	(0.159)	(0.179)	(0.132)	(0.128)	(0.014)
# of Schools/Vill pre-birth	0.439**	-0.390	-0.108	-0.534***	0.003
	(0.198)	(0.238)	(0.173)	(0.182)	(0.012)
Observations	54,789	54,233	54,789	54,233	51,150
R-squared	0.476	0.399	0.563	0.362	0.316
MeanDepVar	7.22	22.93	5.20	18.54	0.13
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
CLUSTER	statedist	statedist	statedist	statedist	statedist
NUM_clusters	528	528	528	528	528

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is the years of education of husbands (or men) in Column (1); age at marriage of husbands in Column (2); years of education of wives in Column (3); age at marriage of wives in Column (4); and, dummy of inter-caste marriage in Column (5). The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table 2: Impact of Schools on Inter-Caste Marriage

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.031**	0.032**	0.032**	0.030**	0.030**
	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
# of Schools/Vill pre-birth	0.009	0.011	0.011	0.010	0.010
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
h_educ_yr		-0.002***	-0.002***	-0.002***	-0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
w_educ_yr			0.000	0.000	0.000
			(0.001)	(0.001)	(0.001)
age_marr_H				0.001	0.001
				(0.001)	(0.001)
age_marr_W					-0.000
					(0.001)
Observations	46,741	46,741	46,741	46,222	46,222
R-squared	0.317	0.317	0.317	0.319	0.319
MeanDepVar	0.13	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502	502

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) is the baseline regression. Column (2) adds husbands' years of education as controls. Column (3) adds wives' years of education; Column (4) and (5) further add age at marriage for husband and wife. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table 3: Impact of schools expansion intensity on years of education, age at marriage and ICM

VARIABLES	(1) h_educ_yr	(2) age_marr_H	(3) w_educ_yr	(4) age_marr_W	(5) icm
Δ in Schools/Vill b.n. 0-9 age (2nd quartile)	0.025	-0.044	-0.016	0.086	-0.003
	(0.090)	(0.095)	(0.070)	(0.070)	(0.007)
Δ in Schools/Vill b.n. 0-9 age (3rd quartile)	0.152	0.210**	0.137	0.171**	0.002
	(0.102)	(0.101)	(0.085)	(0.083)	(0.008)
Δ in Schools/Vill b.n. 0-9 age (4th quartile)	0.399***	0.413***	0.353***	0.277***	0.011
	(0.127)	(0.131)	(0.101)	(0.094)	(0.010)
# of Schools/Vill pre-birth	0.474**	-0.316	-0.023	-0.446**	0.004
	(0.198)	(0.263)	(0.163)	(0.193)	(0.011)
Observations	54,789	54,233	54,789	54,233	51,150
R-squared	0.476	0.399	0.563	0.362	0.316
MeanDepVar	7.22	22.93	5.20	18.54	0.13
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	528	528	528	528	528

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is the years of education of husbands (or men) in Column (1); age at marriage of husbands in Column (2); years of education of wives in Column (3); age at marriage of wives in Column (4); and, dummy of inter-caste marriage in Column (5). The main explanatory variable here is the change in schools (per village) in the first 9 years of age of husbands - split into 4 quartiles, with the first quartile as the base category. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table 4: Impact of Schools on Inter-Caste Marriage for General caste

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.081	0.085*	0.086*	0.088*	0.088*
	(0.050)	(0.049)	(0.048)	(0.049)	(0.049)
# of Schools/Vill pre-birth	0.018	0.021	0.015	0.015	0.013
	(0.039)	(0.038)	(0.038)	(0.038)	(0.038)
h_educ_yr		-0.004*	-0.002	-0.002	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)
w_educ_yr			-0.006***	-0.005***	-0.005**
			(0.002)	(0.002)	(0.002)
age_marr_H				-0.002	0.000
				(0.002)	(0.002)
age_marr_W					-0.004
					(0.003)
Observations	6,517	6,517	6,517	6,474	6,474
R-squared	0.595	0.596	0.598	0.596	0.596
MeanDepVar	0.16	0.16	0.16	0.16	0.16
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	411	411	411	410	410
Caste	General	General	General	General	General

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Further, the sample is restricted to the General caste population. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) is the baseline regression. Column (2) adds husbands' years of education as controls. Column (3) adds wives' years of education; Column (4) and (5) further add age at marriage for husband and wife. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table 5: Impact of Schools on Inter-Caste Marriage for OBC

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.042**	0.042**	0.042**	0.041**	0.042**
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
# of Schools/Vill pre-birth	-0.010	-0.010	-0.010	-0.011	-0.010
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
h_educ_yr		-0.001	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
w_educ_yr			0.000	0.000	0.000
			(0.001)	(0.001)	(0.001)
age_marr_H				-0.000	-0.001
				(0.001)	(0.001)
age_marr_W					0.002
					(0.002)
Observations	19,493	19,493	19,493	19,271	19,271
R-squared	0.412	0.412	0.412	0.413	0.413
MeanDepVar	0.10	0.10	0.10	0.10	0.10
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	486	486	486	486	486
Caste	OBC	OBC	OBC	OBC	OBC

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Further, the sample is restricted to the Other Backward Class (OBC) group population. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) is the baseline regression. Column (2) adds husbands' years of education as controls. Column (3) adds wives' years of education; Column (4) and (5) further add age at marriage for husband and wife. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table 6: Impact of Schools on Inter-Caste Marriage for SC

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.056*	0.057*	0.057*	0.055*	0.055*
	(0.031)	(0.031)	(0.031)	(0.032)	(0.032)
# of Schools/Vill pre-birth	0.041	0.042	0.042	0.051	0.050
	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)
h_educ_yr		-0.001	-0.001	-0.001	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)
w_educ_yr			0.000	0.000	0.000
			(0.002)	(0.002)	(0.002)
age_marr_H				-0.001	-0.001
				(0.001)	(0.002)
age_marr_W					0.001
					(0.002)
Observations	8,150	8,150	8,150	8,047	8,047
R-squared	0.534	0.534	0.534	0.534	0.534
MeanDepVar	0.16	0.16	0.16	0.15	0.15
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	452	452	452	452	452
Caste	SC	SC	SC	SC	SC

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Further, the sample is restricted to the Scheduled Caste population. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) is the baseline regression. Column (2) adds husbands' years of education as controls. Column (3) adds wives' years of education; Column (4) and (5) further add age at marriage for husband and wife. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table 7: Impact of Schools on Inter-Caste Marriage for ST

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	-0.021	-0.022	-0.022	-0.023	-0.023
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
# of Schools/Vill pre-birth	0.022	0.022	0.022	0.024	0.024
	(0.026)	(0.026)	(0.026)	(0.027)	(0.027)
h_educ_yr		-0.004***	-0.004***	-0.004***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
w_educ_yr			0.001	0.001	0.001
			(0.001)	(0.002)	(0.002)
age_marr_H				0.000	-0.000
				(0.001)	(0.002)
age_marr_W					0.001
					(0.002)
Observations	7,272	7,272	7,272	7,122	7,122
R-squared	0.588	0.589	0.589	0.590	0.590
MeanDepVar	0.09	0.09	0.09	0.09	0.09
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	267	267	267	266	266
Caste	ST	ST	ST	ST	ST

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Further, the sample is restricted to the Scheduled Tribe population. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) is the baseline regression. Column (2) adds husbands' years of education as controls. Column (3) adds wives' years of education; Column (4) and (5) further add age at marriage for husband and wife. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

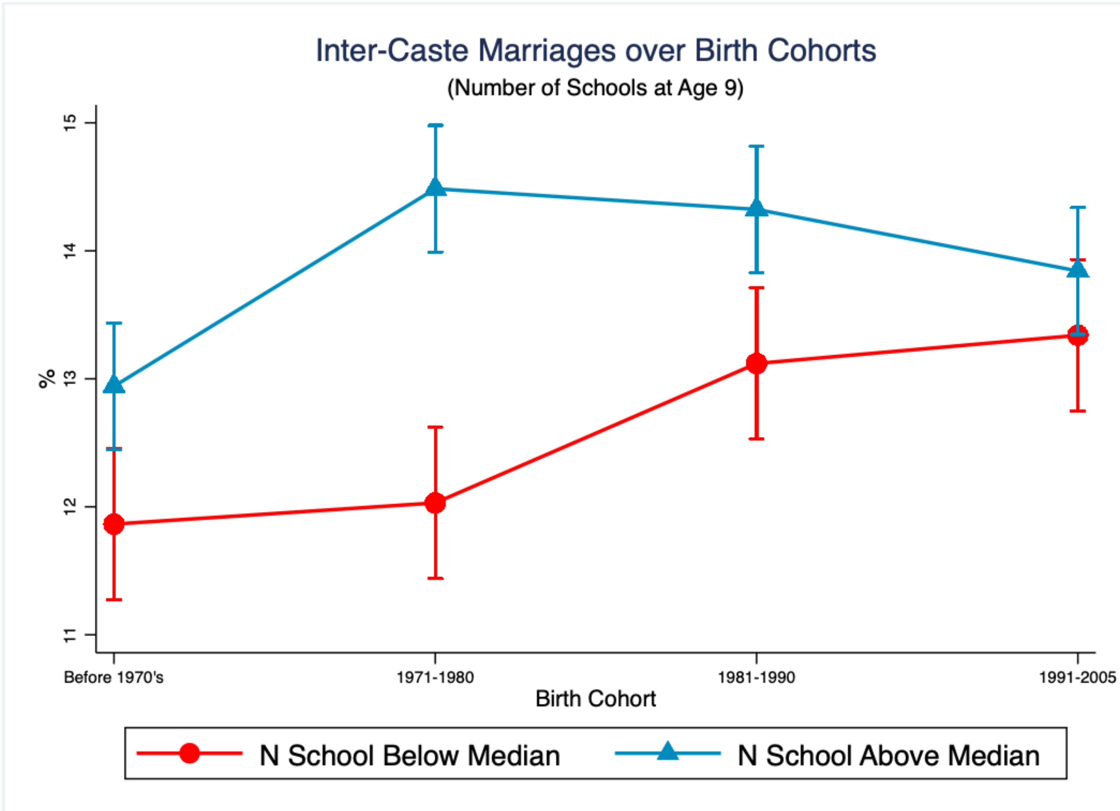
A Appendix Figures

Figure A.I: ICM by husband and wife caste group

Wife's Caste	**Husband's Caste**			
	Scheduled Tribe (ST)	Scheduled Caste (SC)	Other Background Class (OBC)	General Caste (Gen)
Scheduled Tribe (ST)	**87.9 %** (10,584)	**5.8 %** (855)	**1.9 %** (551)	**1.5 %** (178)
Scheduled Caste (SC)	**7.0 %** (836)	**83.3 %** (12,243)	**4.4 %** (1,314)	**3.0 %** (357)
Other Background Class (OBC)	**4.0 %** (476)	**8.8 %** (1,294)	**88.8 %** (26,471)	**12.5 %** (1,503)
General Caste (Gen)	**1.2 %** (141)	**2.1 %** (302)	**5.0 %** (1,481)	**83.1 %** (9,997)

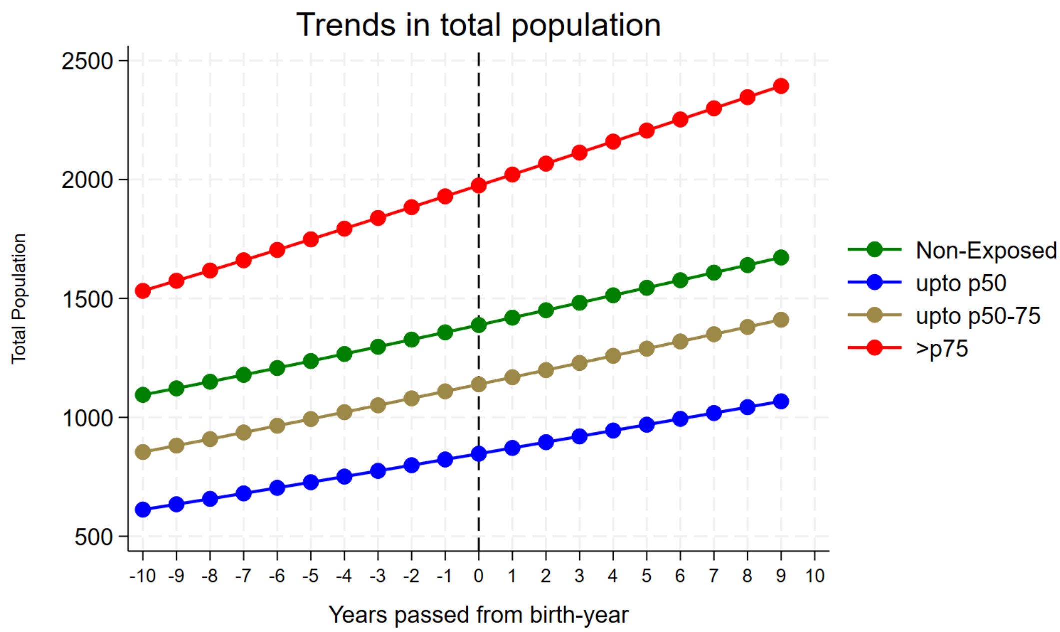
Notes: The table shows the rate of ICM by husband and wife caste categories. For e.g., if we look at the first row, then 87% of the ST marriages occur within the same caste, and out of 13% ICM, 7% is with SC, 4.5% with OBC, and 1.5% with the General caste group.

Figure A.II: ICM by husband and wife caste group



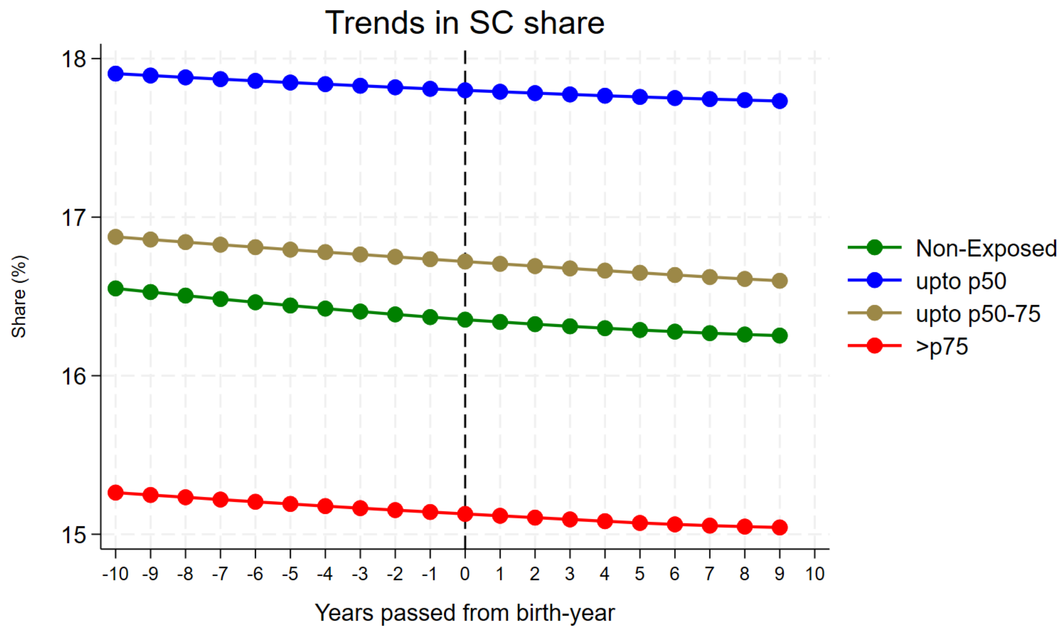
Notes: The figure shows the rate of ICM by two categories- below and above the median, where the median is defined by the total number of schools at the age of 9. It shows that the rate of ICM is higher for all the birth cohort of men who where exposed to more number of primary schools in their school-going age.

Figure A.III: Trends in Total Population



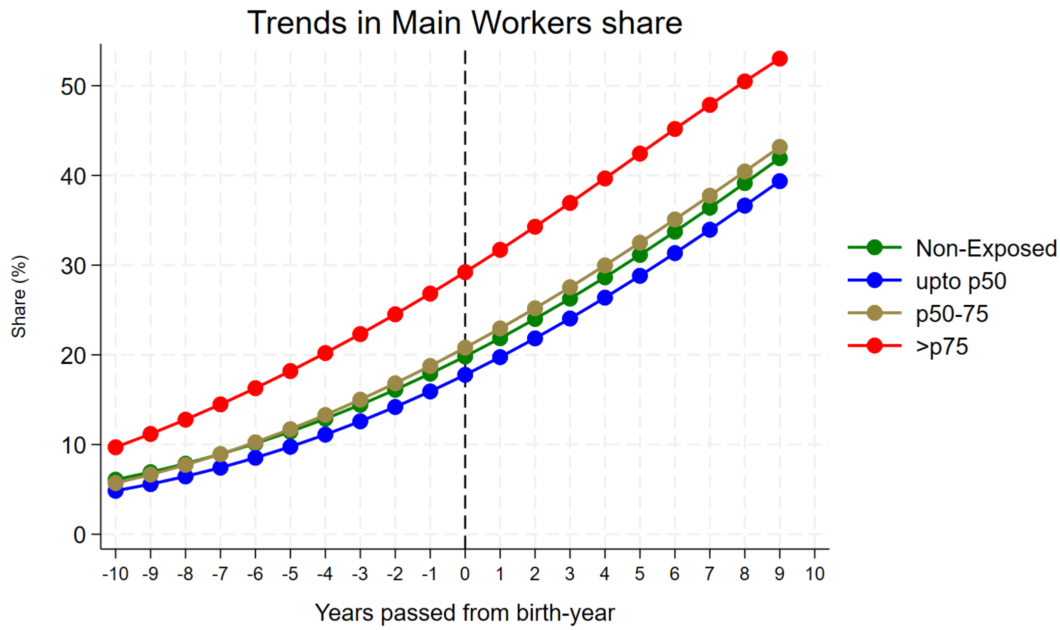
Notes: The figure shows the evolution in the total population in the clusters of exposed and non-exposed populations around their birth years (x=0 is the birth year). The exposure is split into three based on the intensity of exposure. p50 and p75 are the percentile distribution.

Figure A.IV: Trends in SC Population share



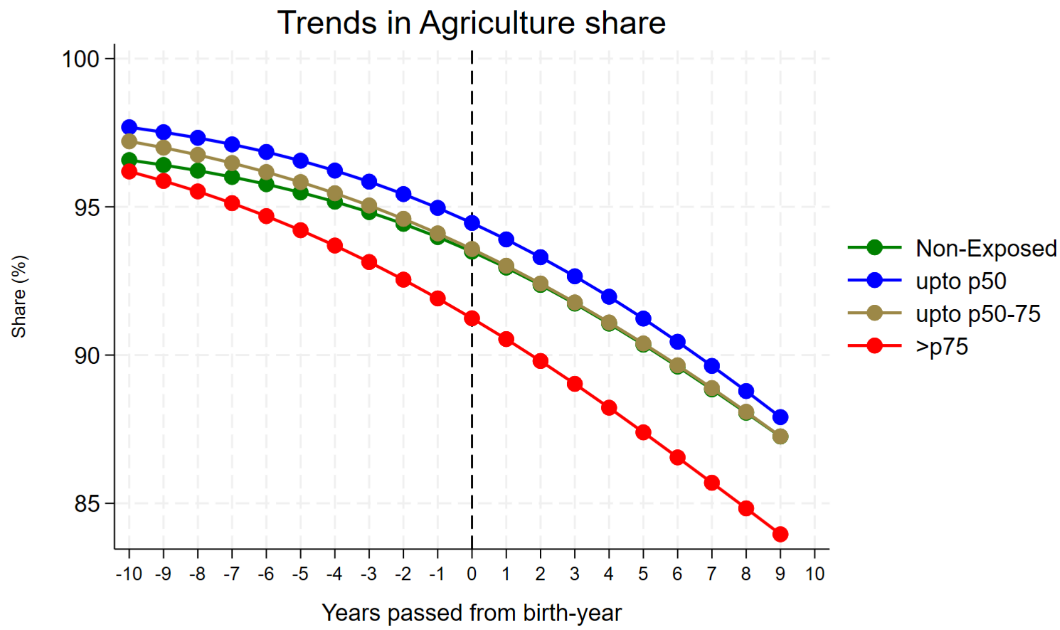
Notes: The figure shows the evolution in the SC population share in the clusters of exposed and non-exposed populations around their birth years (x=0 is the birth year). The exposure is split into three based on the intensity of exposure. p50 and p75 are the percentile distribution.

Figure A.V: Trends in Main Workers Population Share



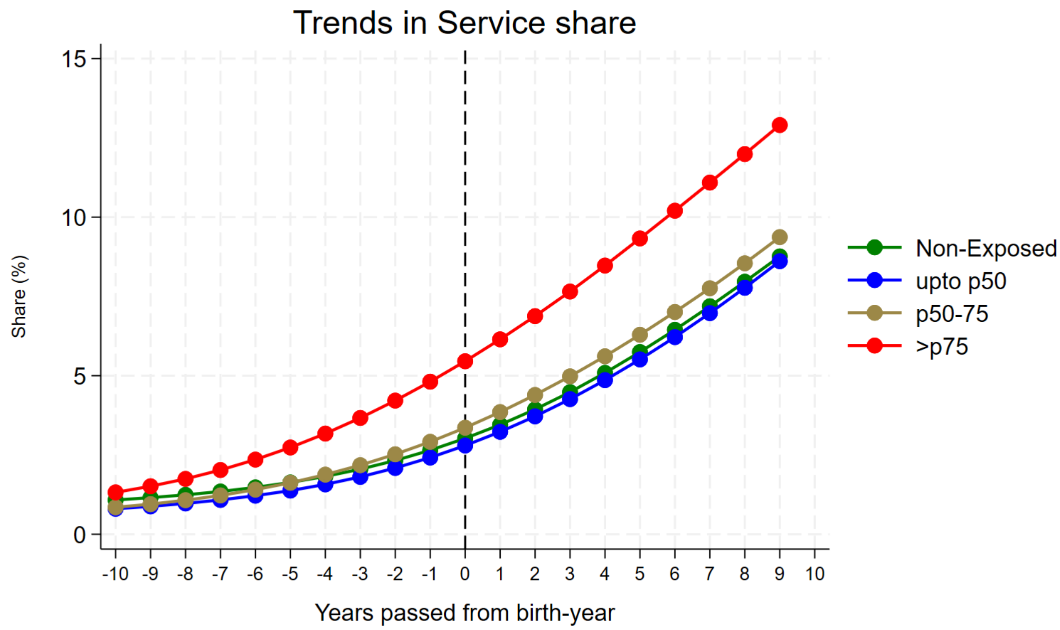
Notes: The figure shows the evolution in the share of the population (male) engaged in full-time working (reporting main workers in the census) in the clusters of exposed and non-exposed populations around their birth years (x=0 is the birth year). The exposure is split into three based on the intensity of exposure. p50 and p75 are the percentile distribution.

Figure A.VI: Trends in Agricultural Share



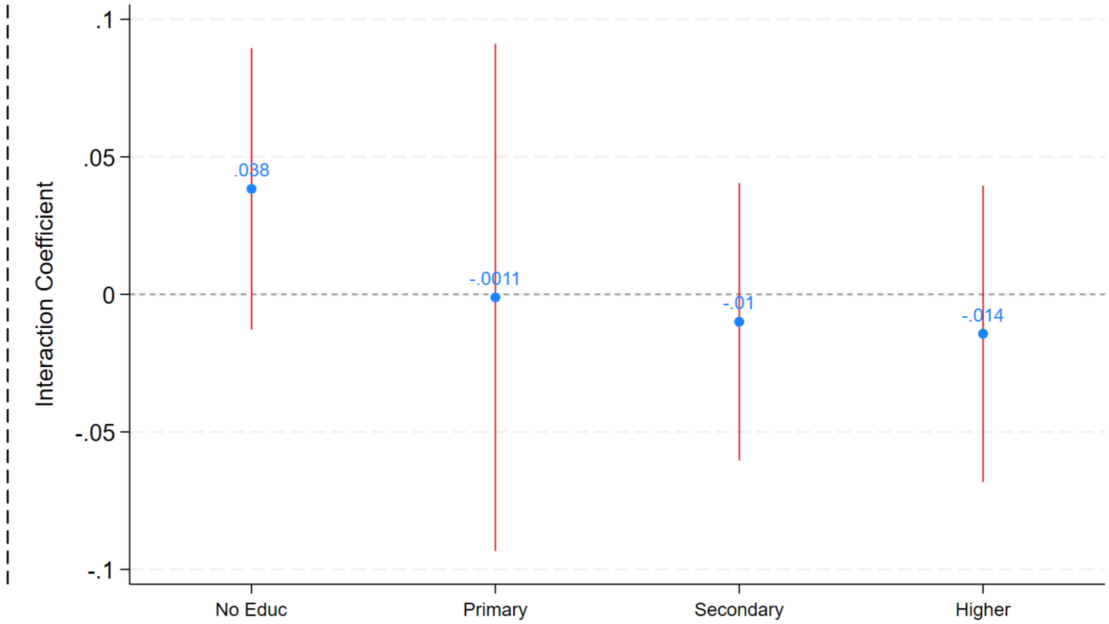
Notes: The figure shows the evolution in the share of the population (male) engaged in the agricultural sector in the clusters of exposed and non-exposed populations around their birth years (x=0 is the birth year). The exposure is split into three based on the intensity of exposure. p50 and p75 are the percentile distribution.

Figure A.VII: Trends in Service Share



Notes: The figure shows the evolution in the share of the population (male) engaged in the service sector in the clusters of exposed and non-exposed populations around their birth years (x=0 is the birth year). The exposure is split into three based on the intensity of exposure. p50 and p75 are the percentile distribution.

Figure A.VIII: Differential Impact of Men's Completed Level of Education



Notes: The figure plot coefficient on the dummy of men's completed education level after the ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever - following Equation 1. The dependent variable is a dummy of inter-caste marriage.

B Appendix Tables

Table B.I: Data Preparation: Matching Rate

States	NFHS (5km buffer zone)	DISE-CENSUS	NFHS & DISE-CENSUS	Unmatched Villages In NFHS but not in DISE-CENSUS	Matching Rate
Andhra Pradesh	6,881	24,334	6,267	614	91%
Assam	19,807	15,532	12966	6,841	65%
Bihar	40,026	29,062	26,289	13,737	66%
Chhattisgarh	12,298	16407	10,423	1,875	85%
Goa	268	205	171	97	64%
Gujarat	11,558	16,697	10,718	840	93%
Haryana	6,140	5,904	5,308	832	86%
Himachal Pradesh	10,552	6,949	4,007	6,545	38%
Jammu & Kashmir	5,419	4,866	4072	1,347	75%
Jharkhand	27,531	22,023	18,690	8,841	68%
Karnataka	17,370	22,577	13,539	3,831	78%
Kerala	923	730	663	260	72%
Madhya Pradesh	36,452	43,052	29,030	7,422	80%
Maharashtra	19,957	36,508	16,964	2,993	85%
Manipur	418	1572	309	109	74%
Odisha	35,667	34297	24,607	11,060	69%
Puducherry	86	82	78	8	91%
Punjab	10,048	9,580	7,738	2,310	77%
Rajasthan	25,124	29,721	17,178	7,946	68%
Sikkim	450	155	155	295	34%
Tamil Nadu	10,642	11,651	7,908	2,734	74%
Tripura	843	775	748	95	89%
Uttar Pradesh	89,252	63,374	53,441	35,811	60%
Uttarakhand	13,877	8,873	7,334	6,543	53%
Total	***401,589***	***404,926***	***278,603***	***122,986***	***69%***

Notes: The table presents the matching rate by state. Column (1) is the name of the state. Column (2) is the number of villages falling within all the 5km buffer zones created. Column (3) is the number of villages in each state coming from the DISE-Census fuzzy matching exercise. Col (4) is the number of villages after matching Col (2) and Col (3). Col (5) is the number of villages unable to be matched. Col (6) is the matching rate. Overall we are able to map 70% of the villages from Col (2).

Table B.II: Simultaneous Change in Other Infrastructure

Simultaneous Δ b.n 0-9 age	(1) Middle School	(2) Secondary School	(3) College	(4) H. Practitioners	(5) Health Centres	(6) Infra index	(7) Roads
Δ in Schools/Vill b.n. 0-9 age	0.077***	0.056***	0.040**	0.293**	0.096***	0.194***	-0.006
	(0.028)	(0.021)	(0.016)	(0.119)	(0.028)	(0.041)	(0.006)
Observations	49,046	49,046	49,046	49,046	49,046	49,046	49,046
R-squared	0.755	0.721	0.984	0.804	0.775	0.786	0.791
MeanDepVar	0.08	0.02	-0.04	-0.06	0.02	0.43	0.08
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502	502	502	502

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The main explanatory variable is the change in schools (per village) in the first 9 years of age of husbands. The dependent variables are changes in other infrastructure in the first 9 years of age: middle schools in Column (1); secondary schools in Column (2); colleges in Column (3); number of medical practitioners in Column (4); number of health centers in Column (5); infrastructure index in Column (6) and availability of pucca roads in Column (7). We also control all infrastructure (per village) one year before birth, household current wealth, and a dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.III: Adding simultaneous change in other infrastructure

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm	(6) icm	(7) icm
Δ in Schools/Vill b.n. 0-9 age	0.030**	0.028*	0.026*	0.025	0.030**	0.034**	0.029**
	(0.014)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)
Δ in Middle school	-0.014						
	(0.020)						
Δ in Secondary school		0.025					
		(0.027)					
Δ in College			0.059				
			(0.042)				
Δ in H. practitioners				0.015			
				(0.016)			
Δ in Health centres					-0.022		
					(0.017)		
Δ in Infra index						-0.014	
						(0.011)	
Δ in Roads							0.015
							(0.028)
Observations	46,741	46,741	46,741	46,741	46,741	46,741	46,741
R-squared	0.316	0.316	0.316	0.316	0.316	0.316	0.316
MeanDepVar	0.13	0.13	0.13	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502	502	502	502

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The main explanatory variable is the change in schools (per village) in the first 9 years of age of husbands. Different columns add the change in various other infrastructures explicitly: middle schools in Column (1); secondary schools in Column (2); colleges in Column (3); number of medical practitioners in Column (4); number of health centers in Column (5); infrastructure index in Column (6) and availability of pucca roads in Column (7). We also control all infrastructure one year before birth, household current wealth, and a dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.IV: Simultaneous Change in Other Infrastructure

Δ b.n 0-9 age	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm	(6) icm	(7) icm
Middle School	-0.007						
	(0.020)						
Secondary School		0.035					
		(0.027)					
College			0.070*				
			(0.041)				
Health Practioners				0.019			
				(0.015)			
Health Centres					-0.013		
					(0.017)		
Infra index						-0.010	
						(0.011)	
Roads							0.015
							(0.028)
Observations	46,741	46,741	46,741	46,741	46,741	46,741	46,741
R-squared	0.316	0.316	0.316	0.316	0.316	0.316	0.316
MeanDepVar	.13	.13	.13	.13	.13	.13	.13
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502	502	502	502

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The main explanatory variable is the change in different infrastructures (per village) in the first 9 years of age of husbands. The infrastructure under consideration is: middle schools in Column (1); secondary schools in Column (2); colleges in Column (3); number of medical practitioners in Column (4); number of health centers in Column (5); infrastructure index in Column (6) and availability of pucca roads in Column (7). We also control all infrastructure (per village) one year before birth, household current wealth, and a dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.V: Simultaneous Change in Other Infrastructure

Simultaneous Δ b.n 0-9 age	(1) Pop	(2) Sex ratio	(3) Child SR	(4) SC share	(5) ST share	(6) Manuf share	(7) Service share	(8) Agri share	(9) Working Pop share
Δ in Schools/Vill b.n. 0-9 age	0.011	0.000	0.000	0.000	0.000	-0.001*	-0.001	0.002	-0.003
	(2.373)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.005)	(0.003)
Observations	49,046	49,046	49,046	49,046	49,046	49,046	49,046	49,046	49,046
R-squared	0.997	0.995	0.995	0.964	0.933	0.979	0.735	0.742	0.897
MeanDepVar	291.57	0.00	0.00	-0.00	-0.00	-0.02	0.06	-0.07	0.22
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502	502	502	502	502	502

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The main explanatory variable is the change in schools (per village) in the first 9 years of age of husbands. The dependent variables are changes in other infrastructure in the first 9 years of age: total population in Column (1); sex ratio in Column (2); child sex ratio in Column (3); share of SC population in Column (4); share of ST population (5); share of male population engaged in manufacturing in Column (6); share of male population engaged in service in Column (7); share of male population engaged in agriculture in Column (8); and total working population share in Column (9). We also control all schools (per village) one year before birth, household current wealth, and a dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.VI: Adding all census controls together

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm
Δ in Schools/Vill b.n. 0-9 age	0.032**	0.031**	0.021	0.031*
	(0.015)	(0.014)	(0.017)	(0.018)
Observations	46,741	46,741	44,749	44,749
R-squared	0.317	0.317	0.321	0.322
MeanDepVar	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502
Adding characs	5 yrs pre-birth	During 0-9 years	5 yrs pre-marriage	All

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is a dummy for movers. Our census controls are demographic (population, share of SC, share of ST, share of female), educational (status of other educational institutions), infrastructure (health, road, communication), and economic (share of agriculture, share of manufacturing, and share of service). In Column (1) controls are census characteristics five years before birth; Column (2) adds the controls during age 0-9; Column (3) adds for five years before the marriage; and Column (4) adds all the controls. We also control all schools (per village) one year before birth, household current wealth, and a dummy for Hindu and Muslim religions. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.VII: Are non-movers different than movers?

VARIABLES	(1) Hindu	(2) Muslim	(3) Others	(4) OBC	(5) General	(6) SC	(7) ST	(8) wealth index
non_mover	0.007	-0.004	-0.003	0.007	-0.010**	0.003	0.001	6,838.170
	(0.006)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.003)	(6,629.041)
Observations	152,001	152,001	152,001	144,637	144,637	144,637	144,637	152,001
R-squared	0.596	0.619	0.581	0.445	0.453	0.372	0.642	0.534
MeanDepVar	0.77	0.11	0.12	0.39	0.18	0.20	0.23	67793.98
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	716	716	716	716	716	716	716	716

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5 men samples living in rural areas. Robust standard errors in parentheses are clustered at the district level. The main explanatory variable is a dummy for non-movers. The dependent variables are dummies for Hindu, Muslim, and Other religions in Columns (1), (2), and (3), respectively. Columns (4)-(8) are dummies for caste groups - Other Backward Caste (OBC), General, SC, and ST in order. Column (8) is the wealth index provided by the survey. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.VIII: Robustness check: Impact of schools on ICM with different subpopulation

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.013	0.040***	0.016	0.017	0.035*
	(0.020)	(0.016)	(0.016)	(0.017)	(0.020)
# of Schools/Vill pre-birth	-0.007	0.001	0.013	0.012	0.015
	(0.017)	(0.014)	(0.014)	(0.015)	(0.016)
Observations	46,857	41,761	46,889	43,491	35,583
R-squared	0.329	0.332	0.327	0.330	0.349
MeanDepVar	0.13	0.14	0.13	0.13	0.14
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	528	525	520	515	513
Condition	W_Abv_22	Excl_ST	Excl_Muslim	Hindu_Only	Hindu_non_ST

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) keeps the wife above 22 years; Column (2) excludes Scheduled Tribe from the sample; Column (3) excludes Muslims; Column (4) keeps hindu only subsample (and no religion dummy controls); and Column (5) is with the sub-sample who are Hindu and not Scheduled Tribe. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.IX: Robustness checks: Impact of schools on ICM with different level of matching

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm
Δ in Schools/Vill b.n. 0-9 age	0.028*	0.025*	0.027*	0.026	0.019
	(0.015)	(0.015)	(0.016)	(0.017)	(0.021)
# of Schools/Vill pre-birth	0.005	0.002	0.000	-0.001	0.003
	(0.013)	(0.013)	(0.014)	(0.015)	(0.015)
Observations	43,992	40,019	32,753	24,056	12,320
R-squared	0.316	0.318	0.320	0.318	0.306
MeanDepVar	0.14	0.14	0.14	0.14	0.14
DHS FE	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist
NUM_clusters	524	514	501	476	373
Condition	$> 50\%$	$> 60\%$	$> 70\%$	$> 80\%$	$> 90\%$

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) keeps where matching rate was at least 50% (i.e. we were able to find the school-level DISE information for more than 50% of the NFHS 5km cluster); Column (2) with matching rate above 60%; Column (3) with matching rate above 70%; Column (4) with matching rate above 80%; and Column (5) with matching rate above 90%. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.X: Differential impact on ICM by education level

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm
Δ in Schools/Vill b.n. 0-9 age	0.028	-0.049	0.023	0.030
	(0.033)	(0.038)	(0.018)	(0.053)
# of Schools/Vill pre-birth	0.084*	-0.014	0.012	0.123
	(0.048)	(0.053)	(0.018)	(0.083)
Observations	7,116	5,665	26,031	2,760
R-squared	0.489	0.514	0.402	0.525
MeanDepVar	0.14	0.13	0.13	0.12
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	443	478	528	407
Condition	no education	primary	secondary	higher

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. The main explanatory variable is the change in schools (per village in the first 9 years of age of husbands. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. Column (1) retains men who have no education; Column (2) is with men who have completed at least primary level education; Column (3) is with men who have completed at least secondary level education; and Column (4) is with men who have completed at least higher level education. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.XI: Placebo Test

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm
Δ in Schools/Vill b.n. 12-17 years	-0.010			
	(0.013)			
Δ in Schools/Vill b.n. 12-18 years		-0.011		
		(0.011)		
Δ in Schools/Vill b.n. 12-19 years			-0.005	
			(0.011)	
Δ in Schools/Vill b.n. 11-19 years				-0.004
				(0.011)
Observations	50,508	49,861	49,139	49,139
R-squared	0.318	0.319	0.321	0.321
MeanDepVar	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	528	528	528	528

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. The main explanatory variables are: the change in schools (per village) for husbands: from 12-17 years in Column (1); from 12-18 years in Column (2); from 12-19 years in Column (3); from 11-19 years in Column (4). All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.XII: Robustness: Different years exposure

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm
Δ in Schools/Vill b.n. 0-7 years	0.019			
	(0.018)			
Δ in Schools/Vill b.n. 0-8 years		0.027**		
		(0.014)		
Δ in Schools/Vill b.n. 0-10 years			0.021	
			(0.014)	
Δ in Schools/Vill b.n. 0-11 years				0.019
				(0.013)
Observations	51,150	51,150	51,150	51,150
R-squared	0.316	0.316	0.316	0.316
MeanDepVar	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	528	528	528	528

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. The main explanatory variables are the change in schools (per village) for husbands: from 0-7 years in Column (1); from 0-8 years in Column (2); from 0-10 years in Column (3); from 0-11 years in Column (4). All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.XIII: Heterogeneity: Public and Private Schools

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm	(6) icm
Δ in Public Schools/Vill b.n. 0-9 age	0.026				0.024	
	(0.018)				(0.018)	
Δ in Private(including aided) Schools/Vill b.n. 0-9 age		0.042***			0.038**	
		(0.016)			(0.016)	
Δ in Public (including aided) Schools/Vill b.n. 0-9 age			0.031*			0.030
			(0.018)			(0.018)
Δ in Private Schools/Vill b.n. 0-9 age				0.026		0.024
				(0.017)		(0.018)
Observations	51,150	51,150	51,150	51,150	51,150	51,150
R-squared	0.316	0.316	0.316	0.316	0.316	0.316
MeanDepVar	0.13	0.13	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	528	528	528	528	528	528
Type	Pub	Pvt(with aided	Pub(with aided	Pvt		

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. The main explanatory variables are the change occurring from 0-9 years for husbands: in public schools (per village) in Column (1); in private schools (per village), including aided schools in Column (2); in public schools (per village), including aided schools (3); in private schools (per village) in Column (4). All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.XIV: Heterogeneity:Low and High Schools Cluster

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm	(5) icm	(6) icm
Δ in Schools/Vill b.n. 0-9 age	0.157	0.178***	-0.037	-0.009	0.099***	0.034
	(0.098)	(0.060)	(0.029)	(0.036)	(0.038)	(0.022)
1.school_vill19_categ#c.change_9_0	-0.163*					
	(0.097)					
2.school_vill19_categ#c.change_9_0	-0.104					
	(0.100)					
3.school_vill19_categ#c.change_9_0	-0.108					
	(0.100)					
4.school_vill19_categ#c.change_9_0	-0.142					
	(0.100)					
Observations	51,150	273	18,017	13,336	9,987	6,725
R-squared	0.317	0.516	0.326	0.400	0.370	0.365
MeanDepVar	0.13	0.14	0.13	0.13	0.14	0.15
DHS FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist	statedist	statedist
NUM_clusters	528	38	383	457	405	281
Criteria		p0	p0-p25	p25-p50	p50-p75	above p75

Notes: The table reports ordinary least square estimations based on the NFHS-4 and 5, married couple-level sample, living in rural areas, husbands above 22 years old, and who have lived in the same place forever. Robust standard errors in parentheses are clustered at the district level. The dependent variable is a dummy of inter-caste marriage. We also control for the number of schools (per village) one year before birth, household current wealth, and dummy for Hindu and Muslim religions. The main explanatory variable is the change in schools (per village) from 0-9 years for husbands in Column(2) and Column(3). In Column (1), the main explanatory variable interacted with a dummy variable for above median schools at birth. All estimations include NFHS 5km buffer, birth year of men, and NFHS round fixed effects.

Table B.XV: Heterogeneity: Share of Scheduled castes

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm
Δ in Schools/Vill b.n. 0-9 age	0.037*	0.046**	0.024	-0.001
	(0.021)	(0.023)	(0.038)	(0.034)
Observations	11,427	12,128	12,941	13,099
R-squared	0.350	0.348	0.337	0.308
MeanDepVar	0.12	0.16	0.14	0.12
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	291	421	420	362
SC_Share_Categ	1st quartile	2nd Quartile	3rd Quartile	4th Quartile

Notes:

Table B.XVI: Heterogeneity: by SC share and total population

VARIABLES	(1) icm	(2) icm	(1) icm	(2) icm
change_9_0	0.055***	0.002	-0.013	-0.076
	(0.019)	(0.030)	(0.043)	(0.056)
Observations	9,534	12,184	12,156	11,227
R-squared	0.362	0.339	0.355	0.331
MeanDepVar	0.16	0.12	0.12	0.13
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	349	352	336	352
SC_share_categ	Low	High	Low	High
Tot_pop	High	High	Low	Low

Notes:

Table B.XVII: Heterogeneity: by village diversity and total population

VARIABLES	(1) icm	(2) icm	(1) icm	(2) icm
Δ in Schools/Vill b.n. 0-9 age	0.055***	0.016	-0.046	-0.023
	(0.019)	(0.035)	(0.038)	(0.054)
Observations	12,427	9,304	9,430	13,918
R-squared	0.354	0.357	0.371	0.322
MeanDepVar	0.15	0.13	0.12	0.13
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	352	377	376	364
Diversity	Low	High	Low	High
Tot_pop	High	High	Low	Low

Notes:

Table B.XVIII: Robustness: adding census controls

VARIABLES	(1) icm	(2) icm	(3) icm	(4) icm
Δ in Schools/Vill b.n. 0-9 age	0.032**	0.031**	0.021	0.031*
	(0.015)	(0.014)	(0.017)	(0.018)
Observations	46,741	46,741	44,749	44,749
R-squared	0.317	0.317	0.321	0.322
MeanDepVar	0.13	0.13	0.13	0.13
DHS FE	Yes	Yes	Yes	Yes
Birth Year FE	yes	yes	yes	yes
Lived Always	yes	yes	yes	yes
Cluster	statedist	statedist	statedist	statedist
NUM_clusters	502	502	502	502
Adding characs	5 yrs pre-birth	During 0-9 years	5 yrs pre-marriage	All

Notes:

References

Aengst, J. (2014). Adolescent Movements: Dating, Elopements, and Youth Policing in Ladakh, India. *Ethnos*, 79(5):630–649.

Allendorf, K. (2013). Going Nuclear? Family Structure and Young Women's Health in India, 1992-2006. *Demography*, 50(3):853–880.

Allendorf, K. and Pandian, R. K. (2016). The Decline of Arranged Marriage? Marital Change and Continuity in India. *Population and Development Review*, 42(3):435–464.

Allport, G. W. (1979). *The Nature of Prejudice*. Basic Books, Inc.

André, P. and Dupraz, Y. (2023). Education and Polygamy: Evidence From Cameroon. *Journal of Development Economics*, 162(February):103068.

Banerjee, A., Duflo, E., Ghatak, M., and Lafortune, J. (2013). Marry for What? Caste and Mate Selection in Modern India. *American Economic Journal: Microeconomics*, 5(2):33–72.

Béteille, A. (1991). The Reproduction of Inequality: Occupation, Caste and Family. *Contributions to Indian Sociology*, 25(1):3–28.

Bhagat, R. (2016). The Practice of Early Marriages among Females in India: Persistence and Change.

Bharadwaj, P., Løken, K. V., and Neilson, C. (2013). Early life health interventions and academic achievement. *American Economic Review*, 103(5):1862–1891.

Bharti, N. K. and Roy, S. (2023). The early origins of judicial stringency in bail decisions: Evidence from early childhood exposure to hindu-muslim riots in india. *Journal of Public Economics*, 221:104846.

Blau, P. M., Beeker, C., and Fitzpatrick, K. M. (1984). Intersecting Social Affiliations and Intermarriage. *Social Forces*, 62(3):585–606.

Boucher, V., Tumen, S., Vlassopoulos, M., Wahba, J., and Zenou, Y. (2020). Ethnic mixing in early childhood: Evidence from a randomized field experiment and a structural model.

Breton, E. (2021). Modernisation, Demographic Change and State-Level Variations in Household Composition in India. *Asian Population Studies*, 17(3):225–249.

Caldwell, J. C., Reddy, P. H., and Caldwell, P. (1983). The Causes of Marriage Change in South India. *Population Studies*, 37(3):343–361.

Chakraborty, K. (2012). Virtual Mate-Seeking in the Urban Slums of Kolkata, India. *South Asian Popular Culture*, 10(2):197–216.

Chou, S. Y., Liu, J. T., Grossman, M., and Joyce, T. (2010). Parental Education and Child Health: Evidence From a Natural Experiment in Taiwan. *American Economic Journal: Applied Economics*, 2(1):33–61.

Chowdhry, P. (2004). Caste Panchayats and the Policing of Marriage in Haryana: Enforcing Kinship and Territorial Exogamy. *Contributions to Indian Sociology*, 38(1-2):1–42.

chun Lin, Y., yih Chen, A. S., and chen Song, Y. (2012). Does Your Intelligence Help to Survive in a Foreign Jungle? The Effects of Cultural Intelligence and Emotional Intelligence on Cross-Cultural Adjustment. *International Journal of Intercultural Relations*, 36(4):541–552.

Corwin, L. A. (1977). Caste, Class and the Love-Marriage: Social Change in India. *Journal of Marriage and the Family*, 39(4):823.

Davis, K. (1941). Intermarriage in Caste Societies. *American Anthropologist*, 43(3):376–395.

Deary, I. J., Batty, G. D., and Gale, C. R. (2008). Bright Children Become Enlightened Adults. *Psychological Science*, 19(1):1–6.

Dommaraju, P. (2010). Instruction Des Femmes et Évolution du Mariage en Inde. *Population*, Vol. 64(4):757–775.

Dommaraju, P. (2016). Divorce and Separation in India. *Population and Development Review*, 42(2):195–223.

Donner, H. (2002). One's Own Marriage: Love Marriages in a Calcutta Neighboourhood. *South Asia Research*, 22(1):79–94.

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

Dyson, T. and Moore, M. (1983). On Kinship Structure, Female Autonomy, and Demographic Behavior in India. *Population and Development Review*, 9(1):35–60.

Esteve, A. and Reher, D. S. (2021). Rising Global Levels of Intergenerational Coresidence Among Young Adults. *Population and Development Review*, 47(3):691–717.

Fuller, C. J. and Narasimhan, H. (2008). Companionate Marriage in India: The Changing Marriage System in a Middle-Class Brahman Subcaste. *Journal of the Royal Anthropological Institute*, 14(4):736–754.

Fuller, C. J. and Narasimhan, H. (2013). Marriage, Education, and Employment Among Tamil Brahman Women in South India, 1891-2010. *Modern Asian Studies*, 47(1):53–84.

Furtado, D. (2012). Human Capital and Interethnic Marriage Decisions. *Economic Inquiry*, 50(1):82–93.

Garrouste, M. and Zaiem, M. (2020). School supply Constraints in Track Choices: A French Study Using High School Openings. *Economics of Education Review*, 78(August):102041.

Ghimire, D. J., Axinn, W. G., Yabiku, S. T., and Thornton, A. (2006). Social Change, Premarital Nonfamily Experience, and Spouse Choice in an Arranged Marriage Society. *American Journal of Sociology*, 111(4):1181–1218.

Golebiowska, E. A. (2007). The Contours and Etiology of Whites' Attitudes Toward Black-White Interracial Marriage. *Journal of Black Studies*, 38(2):268–287.

Goli, S., Singh, D., and Sekher, T. V. (2013). Exploring the Myth of Mixed Marriages in India: Evidence from a Nation-Wide Survey. *Journal of Comparative Family Studies*, 44(2):193–206.

Gundemeda, N. (2020). Caste in Twenty First Century India: Sociological Reflections on University Students' Perceptions in South India. *Asian and African Studies*, 29(1):89–111.

Hathi, P. (2019). Can More Education Reduce Opposition to Intermarriage? Comparing India and the US.

Hodson, G. and Busseri, M. A. (2012). Bright Minds and Dark Attitudes: Lower Cognitive Ability Predicts Greater Prejudice Through Right-Wing Ideology and Low Intergroup Contact. *Psychological Science*, 23(2):187–195.

Jones, G. W. (2017). Changing Marriage. In *Routledge Handbook of Asian Demography*, page 351.

Joshee, R. (2008). Citizenship Education in India: From Colonial Subjugation to Radical Possibilities. In *The Sage Handbook of Education for Citizenship and Democracy*, pages 175–188. Sage Publications, California.

Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. *Annual Review of Sociology*, 24(1):395–421.

Kannan, C. T. (1963). *Intercaste and Inter-Community Marriages in India*. Allied Publishers.

Karve, I. (1968). *Kinship Organization in India*. Asia Pub. House, 3rd edition.

Khanna, G. (2023). Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India. *Journal of Political Economy*, 131(2):549–591.

Mander, H. and Prasad, M. (2014). India Exclusion Report. Technical report, Books for Change, Bangalore, India.

Medhe, R. (2019). Role of Education in Inter-Caste Marriages of Women from Scheduled Castes in India. *Journal of Geographical Studies*, 2(2):79–86.

Mody, P. (2008). *The Intimate State: Love-Marriage and the Law in Delhi*. Routledge.

Munshi, K. (2017). Caste and the Indian Economy.

Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(01):46–98.

Munshi, K. and Rosenzweig, M. R. (2006). Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. *American Economic Review*, 96(4):1225–1252.

Munshi, K. and Rosenzweig, M. R. (2009). Why is Mobility in India so Low? Social Insurance, Inequality, and Growth.

Netting, N. S. (2010). Marital Ideoscapes in 21st-Century India: Creative Combinations of Love and Responsibility. *Journal of Family Issues*, 31(6):707–726.

Prakash, R. and Singh, A. (2014). Who Marries Whom? Changing Mate Selection Preferences in Urban India and Emerging Implications on Social Institutions. *Population Research and Policy Review*, 33(2):205–227.

Qian, Z. and Lichter, D. T. (2007). Social Boundaries and Marital Assimilation: Interpreting Trends in Racial and Ethnic Intermarriage. *American Sociological Review*, 72(1):68–94.

Ray, T., Roy Chaudhuri, A., and Sahai, K. (2020). Whose Education Matters? An Analysis of Inter Caste Marriages in India. *Journal of Economic Behavior and Organization*, 176:619–633.

Ross, A. D. (1961). *The Hindu Family in Its Urban Setting*. University of Toronto Press., Toronto, CA.

Roth, J. and Peck, R. F. (1951). Social Class and Social Mobility Factors Related to Marital Adjustment. *American Sociological Review*, 16(4):478–487.

Rubin, B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75:591–593.

Sarkar, K. (2022). Can Status Exchanges Explain Educational Hypogamy in India? *Demographic Research*, 46(May):809–847.

Sarkar, K. and Rizzi, E. (2020). Love Marriage in India.

Saroja, K. (1999). Intercaste Marriage and Social Dynamics in India: A Critique. *Indian Journal of Social Work*, 60(2):189–192.

Schermerhorn, R. A. (1970). *Comparative Ethnic Relations: A Framework for Theory and Research*. Random House.

Schoon, I., Cheng, H., Gale, C. R., Batty, G. D., and Deary, I. J. (2010). Social Status, Cognitive Ability, and Educational Attainment as Predictors of Liberal Social Attitudes and Political Trust. *Intelligence*, 38(1):144–150.

Schwartz, C. R., Zeng, Z., and Xie, Y. (2016). Marrying Up by Marrying Down: Status Exchange Between Social Origin and Education in the United States. *Sociological Science*, 3:1003–1027.

Sharma, A. (2019). Influence of Parental Education on Nuptial Behavior: An Analysis of Hindu Community in India. *Journal of Population and Social Studies*, 27(2):168–180.

Singh, M., Shekhar, C., and Shri, N. (2023). Patterns in Age at First Marriage and Its Determinants in India: A Historical Perspective of Last 30 Years (1992–2021). *SSM - Population Health*, 22(February):101363.

Song, M. (2009). Is Intermarriage a Good Indicator of Integration? *Journal of Ethnic and Migration Studies*, 35(2):331–348.

Still, C. (2011). The State in the Palli: Dalit Perspectives on the State in Andhra Pradesh. *Contemporary South Asia*, 19(3):315–329.

Thornton, A. (2001). The Developmental Paradigm, Reading History Sideways, and Family Change. *Demography*, 38(4):449–465.

Trautmann, T. R. (1981). *Dravidian Kinship*. Cambridge University Press, Cambridge, UK.

Tritzmann, F.-M. (2011). The Representation of Women and Female Agency in the Indian Online Matrimonial Market. *Internationales Asienforum*, 42(3-4):239–256.

Uberoi, P. (2006). *Freedom and Destiny: Gender, Family, and Popular Culture in India*. Oxford University Press, Oxford, UK.

Vaid, D. (2014). Caste in Contemporary India: Flexibility and Persistence. *Annual Review of Sociology*, 40:391–410.

Wodtke, G. T. (2012). The Impact of Education on Intergroup Attitudes: A Multiracial Analysis. *Social Psychology Quarterly*, 75(1):80–106.