Public School Access and Social Sorting

Evidence from Rural India

Shruti Agarwal[‡]

September 10, 2026

Abstract

Expanding school access reduces travel costs, but by making schooling more local it may also change who attends school with whom. I exploit the nationwide school expansion induced by India's Right to Education Act, which required a public primary school within one kilometre of every habitation, comparing initially underserved villages with already-covered villages before and after implementation. I construct fine-grained local education markets by linking annual records covering nearly one million schools to more than half a million villages, population data at 100-metre resolution, and near-universe records on 1.67 million named rural settlements. These data capture the schools residents can feasibly reach, including nearby schools across village borders. A policy-induced 10-percentage-point increase in access to multiple nearby public schools raises Scheduled Caste–non-Scheduled Caste segregation by 1.67 percentage points, or 4.3 percent of its pre-reform level; the increase is not a mechanical consequence of dividing pupils among more schools. The same expansion also increases the share of non-SC pupils attending schools with no SC pupils, reducing opportunities for cross-caste contact within school. Holding residential composition fixed, a caste-neutral nearest-school benchmark reproduces 85.5 percent of the observed response using cross-village residential differences alone and 98.5 percent when incorporating within-village differentiation. Thus, denser school networks can transmit residential social geography more directly into school composition, weakening the incidental integration created when spatially separated communities share a school. More broadly, the results show how the spatial design of local public provision can shape social integration even when access rules themselves are group-neutral.

Keywords: school choice; social segregation; education policy; Right to Education Act; India
JEL Codes: I21; I24; I28; J15; R23

[*]University of Warwick. Email: shruti.agarwal.1@warwick.ac.uk.

[†]I am especially grateful to Sonia Bhalotra, Sascha Becker, James Fenske, Andreas Stegmann, Devesh Rustagi, and Natalia Zinovyeva for extensive comments and discussions. I also thank Andreas de Barros, Lily Shevchenko, and Sarthak Joshi for helpful comments and discussions. I thank participants at SEEDEC, the Nordic Conference on Development Economics (NCDE), and the Exeter Diversity and Human Capital Workshop, as well as seminar participants at the Warwick Workshop in Experimental Economics and Political Economy, the Warwick Applied Microeconomics Early Stage Seminar, and the Warwick Political Economy and Public Economics Group, for helpful comments and suggestions.

1 Introduction

Expanding the physical reach of schooling has been one of the central education investments of the past half-century. Where schools are sparse, distance raises the time, financial, and safety costs of attendance. Bringing schools closer has increased enrolment and attainment, narrowed gender gaps, and raised learning across a range of low-income settings (Burde and Linden, 2013; Kazianga et al., 2013; Bassi et al., 2026). These effects can persist into adult earnings (Duflo, 2001). Expansions of public provision can also reshape the wider education market by affecting private-school entry, fees, quality, and the allocation of enrolment across sectors (Lucas and Mbiti, 2012; Muralidharan and Sundararaman, 2015a; Neilson et al., 2020; Bau, 2022; Andrabi et al., 2024). But the spatial design of local public provision can affect more than access to a service. It can also determine which populations share the institutions that provide it. For schools, location determines not only whether children can attend, but also whom they attend with. Attendance boundaries, residential sorting, school choice, and the geography of providers all shape school composition (Ferreyra, 2007; Allen et al., 2015; Böhlmark et al., 2016; Monarrez et al., 2022; Monarrez, 2023). Exposure to peers from different social groups can, in turn, affect discrimination, prosocial behaviour, and later outcomes (Rao, 2019a; Lowe, 2021a; Corno et al., 2022a). An expansion of school supply can therefore have competing effects on social mixing. It may bring previously excluded children into shared institutions, but more local provision may also divide children across a larger number of schools whose composition more closely reflects the surrounding communities. The well-established benefits of improved access thus leave open a distinct question: how does a denser school network change who attends school with whom?

I study this question through the neighbourhood-school provisions of India's Right of Children to Free and Compulsory Education Act. The Act, which came into force in 2010, required governments to provide a public primary school within one kilometre of every habitation. A habitation is a named settlement within a village, and a village may contain several of them. The rule was written entirely in terms of distance: it contained no caste criterion, created no binding attendance boundary, and assigned no child to a particular school. I exploit the resulting nationwide expansion by comparing villages that initially fell short of the one-kilometre requirement with villages whose inhabited areas were already covered before implementation. The mandate worked on the dimension it targeted, bringing public schools closer and expanding nearby access, but it also increased segregation between Scheduled Caste (SC) and non-SC pupils across schools within local education markets. Scheduled Castes are communities that have historically faced untouchability and severe social exclusion, and caste continues to shape residential location within and across villages (Munshi and Rosenzweig, 2006a; Munshi, 2019a; Asher et al., 2026). The evidence points to a geographic mechanism: as the public-school network became denser and more local, residential social geography was transmitted more directly into school composition, reducing the incidental integration that can arise when spatially distinct communities share the same school. The finding identifies an additional distributional consequence of expanded access, not the reform's overall welfare effect.

The empirical design compares initially underserved and already-covered villages before and after the Act came into force, absorbing time-invariant differences across villages, district-specific shocks in each year, and differential changes associated with baseline village characteristics.

Eligibility for expansion increased both coverage and the availability of nearby public-school alternatives. In local education markets where multiple recognised primary schools were already operating before the reform, I therefore use the interaction of baseline undercoverage with post-reform implementation as an instrument for the policy-induced increase in the share of residents with at least two public primary schools within one kilometre. I use this variation to estimate the effect of policy-induced increases in nearby public-school access on school segregation.

A central measurement contribution of the paper is to construct local education markets around where residents live and the schools they can feasibly reach. Village boundaries do not determine school attendance, and nearby provision can extend across them: before the reform, the nearest public primary school was located in another village for 44.6 percent of residents. For every inhabited population cell, I identify the public primary schools within one kilometre and aggregate these resident-level school menus using population weights, counting a school whenever it falls within the resident's feasible radius.[1] I build these measures by combining annual administrative records and geocoded locations for nearly one million schools with village boundaries and population data at approximately 100-metre resolution, covering more than half a million rural villages. I supplement these data with independent administrative records on 1.67 million named rural habitations, which reveal social differentiation within villages that is not observable in village-level Census totals. Together, these data allow me to measure both the school network residents could feasibly reach and the fine-scale residential geography through which changes in that network may affect school composition.

The mandate substantially changed the geography of public-school provision. Eligibility increased the share of residents with at least one public primary school within one kilometre by 6.92 percentage points and the share with at least two by 2.71 percentage points. The expansion moved residents primarily from having no nearby public school into one- and two-school environments rather than creating very thick local markets. It also made provision more local: the population-weighted distance to the nearest public primary school fell by 12 percent, and the share of residents whose nearest public school lay across a village boundary fell by 3.8 percentage points. The reform therefore changed not only whether public schooling was nearby, but also the spatial scale over which residents shared nearby provision.

This localisation changed how Scheduled Caste and non-SC pupils were distributed across schools. I measure segregation using the dissimilarity index across the public and private primary schools serving each local education market, which can be interpreted as the share of SC pupils who would need to change schools to equalise the distributions of SC and non-SC pupils. Instrumental-variables estimates imply that a policy-induced 10-percentage-point increase in the share of residents with at least two nearby public schools raises dissimilarity by 0.0167, or 4.3 percent of its pre-reform mean. Equivalently, approximately 17 of every 1,000

[1] Each local education market is centred on the residents of a Census village. I use 'village' and 'local education market' interchangeably where no confusion arises, but the schools serving a market may lie in neighbouring villages. Independent evidence supports allowing the relevant school set to cross this boundary. The NSS 58th Round records the nearest primary school outside the geographical village boundary for 26.4 percent of surveyed villages, while local household–school and school-origin data document actual cross-village attendance (National Sample Survey Organisation, 2003; Afridi, 2011; Programme Evaluation Organisation, Planning Commission, Government of India, 1965). These sources establish the empirical relevance of boundary crossing, not a national percentage of children attending outside their Census village.

SC pupils would need to change schools to offset the increase. This effect is not a mechanical consequence of dividing a finite number of pupils among more or smaller schools: it remains when I subtract the dissimilarity expected under simulated random pupil allocation, while the random component itself does not increase.

The change in allocation is also visible in the school environment experienced by children. In established markets, the same ten-percentage-point access increase raises the share of non-SC pupils attending schools with no SC pupils by 0.66 percentage points, or 3.4 percent of its pre-reform mean. At this margin, the school offers no opportunity for cross-caste interaction among its primary pupils.

The evidence points to a change in geographic pooling. When public schools are sparse, a single school may be the nearest provider for several spatially distinct settlements. If those settlements differ in caste composition, the school can pool their children even in the absence of any formal assignment rule. Expanding the network closer to underserved habitations reduces the area over which each school is the nearest provider and makes provision more local. When residential composition varies over short distances, this localisation can make school composition reflect the surrounding residential geography more closely. To assess whether this mechanism can account for the observed increase in segregation, I construct a caste-neutral geographic benchmark that holds residential composition fixed and assigns every resident to the nearest active public school using the same distance rule for all.

The benchmark reproduces most of the observed response. When villages are treated as internally mixed, so that only differences in social composition across villages affect assignment, the changing public-school network generates 85.5 percent of the estimated increase in segregation. I then incorporate within-village social differentiation at a strength calibrated to the independent habitation data; the benchmark reproduces 98.5 percent of the observed response. This calibration uses neither school locations, policy eligibility, nor observed school segregation. Residential geography combined with a common distance rule is therefore sufficient to reproduce an increase of the observed magnitude.

The geographic benchmark is not a model of observed school choice, and these results do not imply that families ignore school attributes or caste composition. Rather, they establish a more limited but consequential point: an aggregate increase in segregation of the observed magnitude does not require SC and non-SC families to value school caste composition differently. Several additional patterns support this interpretation. Segregation rises within the pre-existing public-school network rather than being driven only by the composition of newly opened schools. I also find little evidence that SC pupils were reallocated towards incumbent schools with worse pre-reform observable inputs, that contemporaneous changes in school inputs account for the estimate, or that selective movement into private schooling explains the response. Taken together, the evidence supports geographic localisation as an explanation capable of accounting for the aggregate increase in segregation, while leaving room for families to respond to school attributes and for other mechanisms to contribute.

The paper contributes first to research on the consequences of expanding local public provision. Existing work on school supply has established that school expansion affects participation, learning, and the responses of public and private providers. I identify a distinct distributional

margin: when the spatial density of a public-service network changes, the mapping from residential geography into the composition of the institutions people use changes with it. In sparse networks, geographically distinct communities may be pooled into common institutions simply because they share the same feasible provider. As provision becomes denser and more local, that incidental pooling weakens and institutional composition comes to reflect residential social geography more closely. The findings suggest a broader implication: the spatial design of public-service networks can shape distributional outcomes even when access rules themselves are group-neutral.

This mechanism also connects the paper to research on school choice, assignment, and segregation. Existing work studies how vouchers, provider entry, jurisdictional fragmentation, attendance zones, and centralised assignment rules affect the distribution of pupils across schools (Ferreyra, 2007; Böhlmark et al., 2016; Monarrez et al., 2022; Monarrez, 2023; Crema, 2026). Here there is no voucher, newly imposed attendance boundary, or formal assignment reform. Instead, the policy changes the density and location of the ordinary network of free public schools. The results therefore identify a distinct margin: changing the density and location of the school network alters how strongly residential geography is passed through into school composition. This differs from cross-sectional decompositions that ask how a fixed configuration of school sites and attendance zones accounts for observed segregation; I identify the causal change in segregation produced when the spatial configuration of provision becomes denser.

Finally, the paper adds causal evidence to research on caste and education in India. Existing work documents fine-scale residential and school segregation and shows that school composition can affect social attitudes (Rao, 2019a; Bailwal and Paul, 2021; Edara, 2025a,c). Research on the RTE Act has instead concentrated on its reservation of places for disadvantaged children in private schools or on aggregate enrolment and learning (Shah and Steinberg, 2019; Romero and Singh, 2022). I study a different part of the same law: the rule governing the spatial reach of public primary provision. The results show that a mandate written entirely in terms of distance can nevertheless reorganise caste composition across schools because it changes the geography of the institutions families can feasibly reach. Evaluations centred only on distance, coverage, or average achievement can therefore miss an important distributional consequence of expanding access: changes in who attends school with whom.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting, Section 3 presents the data and spatial measurement, and Section 4 sets out the empirical strategy. Section 5 reports the effects on access and segregation, Section 6 examines the geographic mechanism and alternative explanations, and Section 7 concludes.

2 Setting

2.1 The neighbourhood-school rule

The Right of Children to Free and Compulsory Education Act (RTE) gave every child aged six to fourteen a right to free and compulsory elementary education in a neighbourhood school.[2]

[2]The Right of Children to Free and Compulsory Education Act, 2009, Act No. 35 of 2009, § 3(1). The Act came into force on 1 April 2010.

The Act did not say how far a neighbourhood school could be. The rules issued under it did. For Classes I–V, governments and local authorities had to provide a primary school within one kilometre's walking distance of the neighbourhood; for Classes VI–VIII, the corresponding distance was three kilometres.[3]

The rule was applied habitation by habitation. Local authorities had to identify which school or schools served each habitation, map schools against habitations, count school-age children each year, and keep records of them.[4] This system was not new. Sarva Shiksha Abhiyan (SSA), the national elementary-education programme running since 2001, already planned provision from the habitation upward, using annual household surveys and DISE school records. The one-kilometre threshold itself was also already operational: under SSA, new primary schools were to be opened in areas where no school existed within one kilometre of a habitation (Ministry of Human Resource Development, Government of India, 2004, p. 125). RTE therefore gave legal force to a distance standard that the school-building programme had already been administering, rather than introducing the threshold for the first time.

In 2002, 87 percent of habitations had a primary school within one kilometre. The remaining 1.61 lakh habitations, roughly 161,000, did not.[5] Coverage continued to expand rapidly under SSA: by 2006–07 the programme reported having opened roughly 1.32 lakh primary schools and 56,000 Education Guarantee Scheme and Alternative and Innovative Education centres (Planning Commission, Government of India, 2008, Chapter 1, pp. 2–3). Thus, although coverage was already wide, a large number of habitations still fell outside the one-kilometre standard.

RTE required nearby provision but did not create a unique assigned neighbourhood school. The Model Rules refer to the neighbourhood school or schools serving a habitation and explicitly allow more than one neighbourhood school where population density warrants it (Ministry of Human Resource Development, Government of India, 2010, Rule 4(5)). Nor was a child required to attend the neighbourhood school: government guidance stated that admission need not be sought only there (Ministry of Human Resource Development, Government of India, 2011, p. 4). Admission was also not confined to the start of the academic year: schools could not refuse late admission, and children had transfer rights in specified circumstances (Government of India, 2009, §§ 5, 15).

The RTE framework separately addressed caste exclusion within schools. Its Model Rules prohibit caste abuse and require that children from disadvantaged groups not be segregated or discriminated against in school activities and common facilities.[6] The neighbourhood-school provision studied here regulated a different margin: proximity to public provision, rather than assignment to integrated public-school catchments.

[3] Ministry of Human Resource Development, Government of India, *Model Rules under the Right of Children to Free and Compulsory Education Act, 2009*, Rule 4(1)(a)–(b), p. 2. States framed their own rules under § 38 of the Act; most followed the central text on the distance norms.

[4] *Model Rules*, Rules 4–6, pp. 2–4. The Rules define school mapping as planning school location to overcome social barriers and geographical distance.

[5] Government of India, *Eleventh Five Year Plan*, Chapter 1, "Education," p. 2, reporting the Seventh All India Educational Survey.

[6] Ministry of Human Resource Development, Government of India (2010, Rules 5(3)–5(4)) prohibit caste, class, religious, or gender abuse and specify that children from disadvantaged groups should not be segregated or discriminated against in classrooms, during mid-day meals, in playgrounds, or in the use of common drinking-water and toilet facilities.

2.2 Caste

Caste remains a major form of social stratification in India. The classical varna scheme describes four broad categories and places Dalit groups outside and below them; in practice, households belong to several thousand endogamous jatis, whose social position varies across place and over time (Dumont, 1980; Gupta, 2000; Deshpande, 2011). Marriage remains overwhelmingly within jati, and caste networks continue to shape access to jobs, credit, and informal insurance (Banerjee et al., 2013; Munshi and Rosenzweig, 2006b; Banerjee and Munshi, 2004; Munshi and Rosenzweig, 2016; Munshi, 2019b).

Administrative data commonly group households into four broad categories: Scheduled Castes (SC), Scheduled Tribes (ST), Other Backward Classes (OBC), and General. SC includes communities that historically faced untouchability and severe social exclusion. ST covers constitutionally recognised tribal communities. OBC is a broader and more heterogeneous category of groups classified as socially and educationally backward. General is a residual category that includes many traditionally upper-caste groups (Deshpande, 2011). SC households make up 16.6 percent of India's population and ST households 8.6 percent (Office of the Registrar General and Census Commissioner, India, 2011).

Caste remains associated with large differences in social and economic outcomes. Practices associated with untouchability persist in parts of rural India.[7] A wider literature documents caste-based differences in employment, credit, and access to public services. Caste also separates households in space. Using approximately 1.5 million neighbourhoods averaging around 125 households, Asher et al. (2026) estimate that 16 percent of SC residents live in neighbourhoods whose population is more than 80 percent SC. They also relate rural segregation to hamlet structure, in which lower-status settlements may lie apart from the main settlement containing schools and other public services. Residential segregation therefore operates within villages and towns, not only between them.

These patterns also appear in schooling. Children grow up in caste-stratified neighbourhoods, and the social composition of schools partly reflects the composition of the settlements they serve (Deshpande, 2011; Asher et al., 2026). School sector also differs by social background: SC and ST children are less likely to attend private schools, while private enrolment is more common among richer and better-educated households (ASER Centre, 2005; Choudhury and Bhattacharya, 2020; Desai and Dubey, 2012).

Caste concentration is not confined to a small tail of schools. Edara (2025b) estimates that 29.1 percent of Indian schools enrolled at least 90 percent of their pupils from a single broad caste category in 2017–18. Close to 60 percent of measured segregation occurred within districts, indicating that caste sorting across schools is substantially a local phenomenon. These national indices are not directly comparable with the village-level measures used below, but they establish both the prevalence and the local character of school segregation. The caste composition of a school can therefore differ substantially from that of the surrounding village.

The empirical analysis focuses on the contrast between SC and non-SC students. SC status

[7] Using the second wave of the India Human Development Survey, Thorat and Joshi (2020) report that approximately 30 percent of rural households acknowledged practising untouchability, compared with approximately 20 percent of urban households. Because the measure is self-reported, it is best interpreted as a lower-bound indicator of the persistence of these practices rather than as a complete prevalence estimate.

identifies a historically disadvantaged caste category that is recorded for enrolled children in the school administrative data. The non-SC group is much more heterogeneous: it includes ST, OBC, and General-caste students, whose social and economic positions differ markedly. We therefore use SC–non-SC as one important margin of caste stratification, not as a complete description of India's caste hierarchy.

2.3 Distance and school choice

Primary-school markets in rural India are local. Most young children travel short distances to school, often on foot, so distance sharply limits the schools a household can practically use. In the NSS 75th Round, 83.3 percent of rural primary pupils travelled to school on foot and 75.3 percent attended a school located less than one kilometre from home (National Statistical Office, 2020, Table 58). The statutory radius therefore describes the margin on which most rural primary-school journeys occur rather than an administrative threshold with little connection to actual attendance patterns.[8] Attendance falls as travel distance rises, particularly for girls, for whom longer journeys can carry additional safety and social costs (Mukhopadhyay and Sahoo, 2016; Singh, 2015; Kumar, 2023).[9]

Government schools nevertheless remain the dominant provider at the primary level in rural India. In 2017–18, 73.7 percent of rural primary students attended government schools, while 5.3 percent attended private aided schools and 20.9 percent private unaided schools (National Statistical Office, 2020, Statement 4.14(R), p. 94). A substantial private sector therefore operates alongside the public network, including in rural areas (Muralidharan and Kremer, 2006; Muralidharan and Sundararaman, 2015b; Chaudhury and Kingdon, 2014). Private enrolment is more common among richer and better-educated households and lower among SC, ST, and Muslim children (Choudhury and Bhattacharya, 2020; Desai and Dubey, 2012; ASER Centre, 2005). Parents who choose private schools pay fees and often travel farther, and descriptive evidence suggests that they distinguish schools by discipline, safety, teacher effort, accountability, and instructional practice (Muralidharan and Sundararaman, 2015b; Mousumi, 2022; Joshi, 2024).

RTE did not require a child to attend the neighbourhood school or the school nearest home. Families could attend another public school, including one outside their village, when it was accessible to them. Village boundaries therefore did not necessarily define a household's schooling options. Distance and availability do not, however, map neatly onto Census-village boundaries. In survey-weighted IHDS-II data, 38.6 percent of households with a rural primary-school pupil report "close to home" as the most important reason for choosing the child's

[8] Revealed behaviour from the Andhra Pradesh voucher experiment also suggests that short distances affect whether a formally available school becomes an effective option. Muralidharan and Sundararaman (2015a) report that having a private school within 0.5 kilometres was associated with a 6.5-percentage-point higher probability of applying for a voucher and, among voucher winners, a 16.8-percentage-point higher probability of accepting it. These are predictive associations rather than randomized effects of distance, and the intervention concerned private-school choice.

[9] Physical distance need not fully characterize the effective feasible set. Using GPS-based distances in rural Pakistan, Jacoby and Mansuri (2015) find that crossing a settlement boundary predicts lower girls' enrolment conditional on distance, with particularly strong constraints when low-caste girls must travel to a school located in a high-caste settlement. The setting and outcome differ from those studied here, but the evidence shows that settlement boundaries can matter beyond distance alone.

school, and 59.4 percent cite either proximity or the absence of another available school. Among government-primary pupils, these shares rise to 45.0 and 70.5 percent, respectively (Desai et al., 2018). Separately, the NSS 58th Round village-facilities survey records the nearest primary school outside the geographical village boundary for 26.4 percent of surveyed villages (National Sample Survey Organisation, 2003). These facts do not imply that households invariably attend the nearest school. They show that geographic proximity is central to the feasible set and that administrative containment is not a reliable substitute for spatial access.[10]

3 Data and measurement

I combine four data sources into a panel with observations by village and year. I use them to define the local primary school market around each village, measure public school access from the places where residents live, and measure the distribution of Scheduled Caste (SC) and non-SC pupils across schools in the market.

I use different sets of schools to measure access and segregation. The RTE distance rule applied to public provision, so the access and eligibility measures use public primary schools only. The segregation measure includes all recognised primary schools, public and private, because households can choose across both sectors.[11]

3.1 Data sources

School records. The District Information System for Education and its successor, the Unified District Information System for Education Plus (DISE/UDISE+), provide annual administrative records for recognised schools in India. The Ministry of Education began collecting these data in 2005. The 2024–25 UDISE+ report, released in 2025, covers 1.47 million schools, 10.12 million teachers, and 246.93 million students nationwide. Each recognised school completes a Data Capture Format with a reference date of 30 September and returns it through the block and district education offices, where officials check and compile the records (Department of School Education and Literacy, 2023). Recognised schools are required to report, and coverage is near universal across management types. Each school has a stable eleven-digit identifier, which allows it to be followed across years. I use records from 2006 to 2018.[12]

I use four parts of these records. First, the opening year records when a school entered local supply. Second, the grades offered identify schools serving primary grades I to V. Third, enrolment by grade, social group, and sex provides SC and non-SC primary enrolment for each school

[10] Actual boundary crossing is also observed directly. Afridi (2011) matches the Census village of residence and the Census village of the public primary school attended in a rural Madhya Pradesh sample; 3.1 percent of pupils attended outside their residence village. A multi-district Planning Commission survey reports that 66 of 132 sampled rural primary schools received pupils from other villages and describes children using an adjoining revenue-village school because it was more convenient (Programme Evaluation Organisation, Planning Commission, Government of India, 1965). These studies establish that cross-village attendance occurs, but neither provides a nationally representative attendance rate.

[11] Recognition is a legal status. Under § 18 of the RTE Act, no school may operate without a certificate of recognition from the appropriate authority. Management categories in the administrative data distinguish government schools, private aided schools, which receive public grants, and private unaided schools, which do not.

[12] Enrolment by caste is too incomplete to use in 2012 and is unavailable in 2013, so the analysis covers the years from 2006 to 2011 and from 2014 to 2018.

and year. These counts form the segregation measures. Fourth, the records report school facilities and staffing. These include classrooms, drinking water, toilets, electricity, boundary walls, teacher counts, qualifications, and appointment type. I use these variables in the mechanism analysis.

Village geography. The unit of analysis is the rural village. I use village polygons to locate residents and to link the panel to Census characteristics. A harmonised national boundary layer assigns each village a polygon and a Census 2011 location identifier (Office of the Registrar General and Census Commissioner, India, 2011). The polygons locate residents within village boundaries. The Census identifiers link each village to population, caste composition, amenities, and other baseline characteristics. The local school market need not stop at the village boundary, as described below.

School locations. The administrative school records report location as text. They give village, block, and district names, but not geographic coordinates. These names are recorded inconsistently across years, are not always unique within districts, and contain no Census village identifier. Linking schools to Census villages has therefore generally required matching administrative and Census place names (Adukia et al., 2020). A match based on names also places a school inside the village whose name it reports. A household near a village boundary, however, may live closer to a school in the neighbouring village.[13]

I therefore locate schools by coordinate rather than by name. I collect latitude and longitude from the Government of India SchoolGIS portal and link them to the administrative records using the school identifier (Department of School Education and Literacy, 2010). I can then measure the distance from a residential location to each school directly, regardless of the village name attached to the school. Appendix A.1 reports match coverage, coordinate checks, and the treatment of duplicate identifiers and conflicting locations.

Population distribution. Residents are not spread evenly within a village, so the village centroid need not represent where they live. I locate residents using the Global Human Settlement Layer population grid at 100-metre resolution, measured around 2010 (Schiavina et al., 2022). Cells with positive recorded population count as inhabited, and their populations serve as weights when I aggregate access to the village level. I use the same population surface in every year, so measured access changes with the school stock rather than with changes in the population weights.

3.2 Building the village panel

The four sources combine through school and village identifiers. School identifiers link the administrative records to geographic coordinates. The coordinates place schools relative to

[13]This measurement problem appears directly in other Indian education surveys. In the early rounds of the India Early Childhood Education Impact Study, institution visits were restricted to institutions inside the sampled village. The study notes that this prevented institution-level verification for children who attended outside their village (Kaul et al., 2017). Restricting school observation to the sampled village can therefore omit institutions actually used by sampled residents.

village polygons, and Census identifiers link the polygons to village characteristics. The resulting panel is unbalanced because records of enrolment and caste composition are not complete for every village in every year.

Measuring access also requires knowing which public primary schools existed in each year. I use the recorded primary opening year where available and the school's general opening year otherwise. Schools known to predate the panel but missing both fields enter in 2006, the first panel year. I drop schools with no usable information on opening year.

Let o_s denote the entry year for school s. The stock of public primary schools in year t is

$$\mathcal{A}_t = \{s : s \text{ is publicly managed, offers primary grades, has a valid location, and } o_s \leq t\}. \quad (1)$$

A school enters $\mathcal{A}_t$ in its entry year and remains thereafter. A gap in annual reporting therefore does not remove it from the access measures. Appendix A reports the sources of information on opening year and results under alternative rules.

3.3 Measurement

3.3.1 The local school market

A household's options are not confined to schools located inside its own village. Pupil residential locations are not observed in the administrative school records, however, so I cannot reconstruct a separate feasible menu for each enrolled child. I therefore define the local education market for village v in year t spatially around the village polygon P_v. It contains recognised schools with positive primary enrolment that lie inside or within one kilometre of the polygon:

$$\mathcal{S}_{vt} = \left\{ s : \begin{array}{l} s \text{ is a recognised school with positive primary enrolment in year } t, \\ \text{dist}(s, P_v) \leq 1{,}000 \end{array} \right\}. \quad (2)$$

This polygon-centred market and the resident-centred access measure defined below have different spatial anchors because they use different information. The access measure is constructed from inhabited population cells, while the segregation market must be constructed without pupil residence coordinates. Both avoid treating the Census-village boundary as a binding school-market boundary.

Schools inside the polygon have distance zero. The market also includes schools of any management type within one kilometre outside the boundary. A school near a boundary can therefore belong to the markets of neighbouring villages. Unlike the cumulative stock in equation (1), $\mathcal{S}_{vt}$ uses annual enrolment records because caste composition must be observed in each year.

3.3.2 Access from where residents live

The market states which schools are near a village. Access states how many are near its residents. Let $\mathcal{C}_v$ denote the inhabited GHSL cells assigned to village v and p_c the population of cell c.

For each cell and year, I count the public primary schools within one kilometre:

$$N_{ct} = \sum_{s \in \mathcal{A}_t} 1\{\text{dist}(c, s) \leq 1{,}000\}\,. \tag{3}$$

A nearby school counts regardless of which village contains it. Distances are measured in metres in a common projected coordinate system. They represent geographic proximity rather than travel time.

I aggregate these counts using the fixed population weights. For $k \geq 1$,

$$A_{vt}^{(k)} = \frac{\sum_{c \in \mathcal{C}_v} p_c \, 1\{N_{ct} \geq k\}}{\sum_{c \in \mathcal{C}_v} p_c}. \tag{4}$$

$A_{vt}^{(1)}$ is the share of village residents with at least one public primary school within one kilometre. $A_{vt}^{(2)}$ is the share with at least two. I also construct $A_{vt}^{(k)}$ for $k = 3, 4, 5$ and the exact menu shares

$$A_{vt}^{(=j)} = \frac{\sum_{c \in \mathcal{C}_v} p_c \, 1\{N_{ct} = j\}}{\sum_{c \in \mathcal{C}_v} p_c}, \tag{5}$$

which record the shares of residents with exactly j public primary schools within one kilometre.

In the IV specifications, I scale the main endogenous variable as $X_{vt} = A_{vt}^{(2)}/0.10$. A one-unit increase in X_{vt} therefore represents an increase of ten percentage points in the share of residents with at least two public primary schools within one kilometre. The main measures use the statutory radius of one kilometre. The appendix reports alternative radii.

3.3.3 Segregation

Let SC_{st} and $NONSC_{st}$ denote SC and non-SC primary enrolment in school s. Let SC_{vt} and $NONSC_{vt}$ denote the corresponding totals over $\mathcal{S}_{vt}$. I measure segregation with the dissimilarity index (Duncan and Duncan, 1955):

$$D_{vt}^{\text{obs}} = \frac{1}{2} \sum_{s \in \mathcal{S}_{vt}} \left| \frac{SC_{st}}{SC_{vt}} - \frac{NONSC_{st}}{NONSC_{vt}} \right|. \tag{6}$$

The index runs from zero to one. It is zero when SC and non-SC pupils have the same distribution across schools. It rises as the groups concentrate in different schools. Its value is the share of either group that would have to change schools for the two distributions to coincide.

Dissimilarity measured on small units is upward biased. Randomly allocating a finite number of pupils across schools produces positive dissimilarity even when neither group is systematically concentrated, and the bias increases as schools become smaller. Cortese et al. (1976) documented this problem, and Carrington and Troske (1997) showed that standard indices therefore combine random allocation with systematic segregation. I follow their approach and compare observed dissimilarity with the level expected under random allocation.[14]

[14] Other adjustments are available. Rathelot (2012) models the unit-level minority probability as a mixture of Beta distributions. d'Haultfœuille and Rathelot (2017) obtain sharp bounds under a nonparametric binomial mixture. Allen et al. (2015) propose model based adjustments for schools in England. I use simulation because it conditions on the realised number and size of schools in each market and produces observed, expected random, and beyond random dissimilarity on the same sample.

For each village and year, I hold fixed the realised set of schools, each school's caste-valid enrolment, and the market totals of SC and non-SC pupils.[15] I then allocate the fixed number of SC pupils at random and without replacement across the realised school places and recalculate the index. Repeating this procedure 100 times gives expected dissimilarity under random allocation,

$$D_{vt}^{\text{rand}} = \frac{1}{100} \sum_{r=1}^{100} D_{vt}^{(r)}, \tag{7}$$

and I define dissimilarity beyond random allocation as

$$D_{vt}^{\text{beyond}} = D_{vt}^{\text{obs}} - D_{vt}^{\text{rand}}. \tag{8}$$

Only caste allocation varies across draws. The realised schools, school sizes, and market totals by caste stay fixed.

3.4 Construction of baseline eligibility

I define baseline eligibility using the one-kilometre access standard established by the RTE framework. A village is eligible if, in 2008, at least one inhabited location lies more than one kilometre from the nearest public primary school. I construct the measure spatially from inhabited population locations and the pre-reform stock of public primary schools. Eligibility is therefore determined by the least-served inhabited location in the village: an eligible village may still have a large share of its residents living within one kilometre of a public primary school.

I use 2008 as the baseline year, before the transition into RTE implementation, and hold eligibility fixed throughout the panel. Among the 339,858 villages in the 2008 established-market summary sample, 56.3 percent are eligible. Consistent with the construction of the measure, all residents of ineligible villages have a public primary school within one kilometre at baseline, while the corresponding share is 75.1 percent in eligible villages. Appendix A.3 describes the spatial construction of the eligibility measure in detail.

3.5 Analysis samples and summary statistics

I use two analysis samples. The first is the broad policy sample used to measure access and local market formation. A village enters if the Census classifies it as rural, it has a valid polygon and positive gridded population, and its baseline eligibility classification is non-missing. Membership does not depend on whether segregation can be measured. The sample contains 524,639 villages and 4,769,014 observations by village and year.

The second is the established market sample used to study caste allocation. A village enters if its local market contained at least two active recognised primary schools, public or private, in both 2006 and 2007. Membership depends only on market structure before the reform. Neither the 2008 school count nor later segregation outcomes affect it. This restriction excludes markets

[15] Caste-valid enrolment is primary enrolment in schools reporting non-missing SC and non-SC counts that sum to reported total primary enrolment. Schools failing this check are excluded from the market for that year.

with only one school, where dissimilarity is mechanically zero and the arrival of a second school would raise measured segregation by construction.

Within the established market sample, an observation enters estimation when the access measure, baseline covariates, and all three segregation measures are available. These are observed dissimilarity, expected dissimilarity under random allocation, and their difference. The estimation sample contains 3,185,709 observations from 350,331 villages in 554 districts. The first stage, reduced form, and two-stage least squares estimates use the same observations.

Descriptive statistics Table 1 describes the broad policy sample and the established-market sample at baseline. In the established-market sample, 86.0 percent of residents had at least one public primary school within one kilometre in 2008 and 50.2 percent had at least two. The average local market contained 5.65 recognised primary schools, of which 4.96 were public and 0.68 private. These villages therefore already contained multiple schools before the reform; the segregation analysis does not rely on the creation of a multi-school market for the first time.

Eligible and ineligible villages differ sharply in the geography and public provision that determine exposure to the one-kilometre rule. Eligible villages average 5.64 square kilometres, compared with 1.77 among ineligible villages, and have roughly twice the population, 1,882 against 961. They nevertheless have fewer nearby public primary schools, 4.15 compared with 6.01. In 2008, 75.1 percent of their residents had at least one public primary school within one kilometre and 33.8 percent had at least two, compared with 100 and 71.3 percent among ineligible villages. The number of nearby private schools, by contrast, is the same in the two groups, at 0.68.

Eligibility is not otherwise balanced across villages. Eligible villages have greater road, bus, mobile, and digital connectivity. SC population shares, however, are nearly identical, at 19.6 percent in eligible villages and 19.7 percent in ineligible villages, while ST shares are 13.7 and 12.6 percent. The one-kilometre norm therefore binds especially in larger villages with thinner public-school coverage, rather than in villages with systematically higher baseline SC shares. These are descriptive level differences; the empirical design in Section 4 uses changes over time within villages rather than baseline equality between eligible and ineligible places.

Panel C reports baseline segregation in the established-market sample. Mean observed dissimilarity between SC and non-SC pupils is 0.411. Of this, 0.113 is the dissimilarity expected under random allocation given the realised number and sizes of schools, leaving mean beyond-random dissimilarity of 0.298. Eligible villages begin with somewhat lower observed dissimilarity than ineligible villages, 0.391 compared with 0.438. The corresponding beyond-random means are 0.282 and 0.319. The analysis below asks whether these pre-existing level differences are followed by differential changes after implementation of the RTE access mandate.

4 Empirical strategy

I use eligibility under the RTE one-kilometre norm to estimate how an expansion in nearby public-school access affected the allocation of SC and non-SC pupils across primary schools. The empirical analysis follows the policy chain. I first estimate whether eligibility changed the

number of public primary schools that residents could reach within one kilometre. I then use the eligibility-induced increase in access to at least two public schools to estimate the effect of expanded feasible choice on caste segregation. This section presents the estimating equations, describes the variation used for identification, and discusses the evidence on the first stage, counterfactual trends, spatial spillovers, exclusion, and monotonicity.

4.1 Design and estimating equations

The outcome is caste segregation across the recognised primary schools in a village's local market. The main measure is observed dissimilarity between SC and non-SC pupils. I also report expected dissimilarity under random allocation and the difference between observed and expected dissimilarity, as defined in Section 3.3.3. The treatment is realised access to multiple nearby public schools, $X_{vt} = A_{vt}^{(2)}/0.10$, so that one unit represents a ten-percentage-point increase in the share of residents with at least two public primary schools within one kilometre (Section 3.3.2). The relationship of interest is

$$Y_{vt} = \alpha_v + \gamma_{d(v),t} + \beta X_{vt} + \mathbf{B}'_{v0}\Theta_t + \varepsilon_{vt}, \tag{9}$$

where α_v denotes village fixed effects, $\gamma_{d(v),t}$ denotes district-by-year fixed effects, and $\mathbf{B}'_{v0}\Theta_t$ allows outcome paths to vary with observed village characteristics. OLS estimation of equation (9) need not recover the effect of expanded school access. New schools may be placed where population, enrolment demand, settlement patterns, or administrative priorities are also changing. These factors may independently affect the allocation of pupils across schools. I therefore instrument realised multiple-school access using eligibility under the one-kilometre norm in 2008.

As described in Section 3.4, a village is eligible if at least one inhabited location was more than one kilometre from the nearest public primary school in 2008. The RTE Act was enacted in August 2009 and came into force on 1 April 2010. Defining $Post_t = 1\{t \geq 2010\}$, the instrument is

$$Z_{vt} = Eligible_v \times Post_t.$$

The pooled specifications exclude 2009 because it falls between enactment and implementation. The event studies retain 2009 and report it separately as a transition year. All eligible villages share the same implementation date, so the design does not rely on comparisons across staggered treatment cohorts.

Eligibility is fixed using the pre-reform school-access environment. It is not randomly distributed across villages: the one-kilometre norm is more likely to bind in geographically larger and less-served places. In the established-market sample, eligible villages have an average area of 5.6 square kilometres, compared with 1.8 among ineligible villages, and average multiple-school access of 0.338, compared with 0.713. SC population shares are similar across the two groups, at 0.196 and 0.197. Table 1 reports the remaining comparisons. These level differences motivate the use of village fixed effects and a design based on changes over time. The identifying assumption concerns the evolution of eligible and ineligible villages in the absence of the reform, rather than equality of their baseline levels.

The one-kilometre norm directly concerns access to at least one public primary school. The segregation outcome, however, is defined over the allocation of pupils across schools, making access to multiple schools the relevant treatment margin. A school added to serve an uncovered habitation may also give residents who previously had one nearby public school access to a second. Figure 1 illustrates this resident-centred calculation. Because the one-kilometre radius is drawn around inhabited locations, both access and the relevant school market may cross the administrative boundary of the village. The same construction is used throughout the analysis.

I begin by estimating how eligibility changed the local menu of public schools in the broad policy sample defined in Section 3.5. For access outcome Q_{vt}, alternately access to at least one public school, access to at least two public schools, or the share of residents with exactly j nearby public schools, I estimate

$$Q_{vt} = \alpha_v + \gamma_{d(v),t} + \theta Z_{vt} + \mathbf{B}'_{v0}\Delta_t + u_{vt}. \tag{10}$$

Table 2 reports the resulting changes in first access, multiple-school access, and the distribution of public-school menus. Reporting the entire menu is useful because the policy may move residents across several access categories. The exact-two share is not itself a monotone measure of access: a move from two nearby schools to three increases access to at least two schools while reducing the exact-two share. The first-stage treatment therefore remains the cumulative share with access to at least two schools.

The segregation analysis uses the established-market sample defined in Section 3.5. This sample consists of villages whose local market contained at least two active recognised primary schools, public or private, in both 2006 and 2007. The restriction focuses the analysis on markets in which multiple schools were already operating before the reform. It therefore separates changes in the allocation of pupils across an existing set of schools from changes associated solely with the initial appearance of a multi-school market. Sample membership is determined entirely by market structure in 2006 and 2007; it does not use eligibility in 2008, subsequent school entry, or segregation outcomes.

In this sample, I estimate the first stage and reduced form:

$$X_{vt} = \alpha_v + \gamma_{d(v),t} + \pi Z_{vt} + \mathbf{B}'_{v0}\Phi_t + u_{vt}, \tag{11}$$

$$Y_{vt} = \alpha_v + \gamma_{d(v),t} + \tau Z_{vt} + \mathbf{B}'_{v0}\Lambda_t + e_{vt}. \tag{12}$$

I then estimate equation (9) by two-stage least squares, using Z_{vt} as the instrument for X_{vt}. The first-stage, reduced-form, and 2SLS estimates use identical observations. Observed dissimilarity, D_{vt}^{obs}, is the main outcome. Expected dissimilarity under random allocation, D_{vt}^{rand}, holds fixed the realised schools, their enrolment totals, and the market-level caste totals. It captures the finite-enrolment unevenness associated with the observed market structure. The difference $D_{vt}^{\text{beyond}} = D_{vt}^{\text{obs}} - D_{vt}^{\text{rand}}$ measures dissimilarity beyond this benchmark. Estimating the same reduced form for all three outcomes shows whether the policy-induced change in observed dissimilarity is accompanied by a corresponding change in the random-allocation component. Table 3 reports these estimates; Appendix Table A3 reports Theil's entropy index and SC isolation.

All specifications include village fixed effects and district-by-year fixed effects. The former absorb time-invariant differences across villages, including baseline geography and historical school provision. The latter absorb shocks common to villages in the same district in a given year, including district-level implementation, administrative changes, and common changes in schooling conditions. I also interact year indicators with village characteristics recorded in the 2011 Census: log population, log area, the SC and ST population shares, all-weather road access, bus access, mobile coverage, digital connectivity, a public distribution shop, and an ATM. These interactions allow the time path of the outcome to differ flexibly with observed village characteristics.

The Census characteristics were recorded after RTE implementation began, so I examine directly whether their inclusion matters. Appendix Table A5 repeats the first stage, reduced form, and 2SLS estimates on the identical sample after removing all Census-2011 interactions. Section 5.5 reports the resulting estimates, which are nearly unchanged. This audit therefore checks that the identifying comparison is not driven by conditioning on the Census-2011 characteristics.

Standard errors are clustered by district, allowing errors to be correlated across villages and years within a district. The main pooled sample contains 350,331 villages, 554 district clusters, and 3,185,709 village-year observations. It covers 2006–2011 and 2014–2018, with 2009 excluded from the pooled specifications. I do not condition on contemporaneous school supply, enrolment, private-school activity, or school inputs because each may respond to the expansion in access (Angrist and Pischke, 2009). Section 6 studies these responses as mechanisms.

4.2 First stage and policy timing

Instrument relevance requires eligibility to change access to at least two nearby public primary schools. Panel A of Table 3 reports the first stage in the established-market sample. Eligibility increases multiple-school access by 3.25 percentage points. Relative to the eligible group's 2008 mean of 0.338, this is a 9.6 percent increase. The first-stage F statistic is 35.9. The magnitude is therefore both economically meaningful and sufficiently precise for the baseline 2SLS analysis.

The annual access estimates provide additional evidence on when this first stage emerges. Appendix Figure A3 reports the access event study with 2008 omitted. Before 2008, access to at least two schools was declining in eligible villages relative to ineligible villages; after enactment and implementation, the series reverses direction. Figure A4 displays this break in the access path, and Section 5.4 reports the annual estimates.

This pattern changes how the first stage should be interpreted. A comparison based on continuation of the pre-reform path would predict a further relative decline in multiple-school access among eligible villages. The observed post-reform increase therefore represents a break from, rather than a continuation of, the earlier differential trend. Appendix Figure A5 formalises this comparison by extending the estimated pre-reform path forward and allowing for estimation uncertainty. Under exact continuation of the measured pre-reform trend, access would have continued to decline in eligible villages relative to controls. The observed reversal is therefore inconsistent with that particular counterfactual, but does not by itself establish that the pooled first stage is free of differential-trend bias.

I also add an eligible-specific linear calendar-time trend directly to the pooled first-stage regression. This is a demanding sensitivity check because a full-period trend can absorb part of a gradual policy response as well as any differential path that would otherwise have continued. Appendix Table A7 reports the specification, and Section 5.4 discusses the resulting first stage together with the sorting estimates.

4.3 Identification and supporting evidence

The reduced-form coefficient τ compares the change in segregation after RTE implementation between eligible and ineligible villages within the same district and year. Its causal interpretation requires eligible and ineligible villages to have followed comparable counterfactual paths in the absence of the reform, conditional on the fixed effects and controls. The 2SLS interpretation additionally requires eligibility to affect segregation through the instrumented change in multiple-school access and requires eligibility to weakly increase potential access to at least two schools. Because the school markets are resident-centred and may cross village boundaries, I also examine whether nearby ineligible villages share the policy-induced expansion. I discuss each issue in turn.

4.3.1 Counterfactual trends in caste sorting

I first examine the evolution of caste sorting around the reform using the event-study specification

$$Y_{vt} = \alpha_v + \gamma_{d(v),t} + \sum_{k \neq 2008} \delta_k \, \textit{Eligible}_v \, 1\{t = k\} + \mathbf{B}'_{v0}\Psi_t + e_{vt}. \tag{13}$$

I use 2008 as the omitted year because eligibility is measured from the school-access environment in that year. This choice also avoids using 2009, the period between enactment and implementation, as the reference category. The 2009 coefficient is instead shown explicitly and may capture anticipation or early implementation. Figure 2 presents the main dynamics, and Appendix Figure A6 reports the complete 2008-base event study for observed, random-allocation, and beyond-random dissimilarity.

The pre-reform coefficients provide a first diagnostic of the counterfactual-trends assumption. I examine both joint tests of the 2006 and 2007 coefficients and the corresponding linear slopes for observed, random-allocation, and beyond-random dissimilarity. Section 5.4 reports these diagnostics together with the post-reform dynamics.

The event-study coefficients are informative about both the presence and the direction of the pre-reform difference. The estimated pretrend is negative, whereas the post-reform coefficients are positive. If the earlier path had simply continued, eligible villages would have experienced lower relative segregation after 2008. Continuation of the measured pretrend would therefore work against the positive post-reform change. At the same time, the short pre-period means that a conventional pretrend test alone is not sufficient to characterise the range of plausible counterfactual paths. I therefore complement the event study with three exercises that impose progressively different restrictions on departures from parallel trends.

The first exercise adds eligibility interacted with a linear calendar-time trend to the pooled first-stage, reduced-form, and 2SLS equations. This allows eligible villages to remain on a

different linear path throughout the estimation period. Appendix Table A7 reports how the adjustment changes the policy reduced form, the first stage, and the 2SLS scaling estimate.

The second exercise asks what the early post-reform estimates would look like if the measured pre-reform path continued exactly. This is the most direct extrapolation of the pre-period movement. The exercise treats the pre-reform linear path as the counterfactual and incorporates uncertainty in its estimation (Appendix Figure A7).

The third exercise uses the relative-magnitude restrictions proposed by Rambachan and Roth (2023). Rather than imposing exact continuation, this approach allows the post-reform violation of parallel trends to differ from the pre-reform path. The sensitivity parameter $\bar{M}$ limits the magnitude of that violation relative to the largest observed pre-reform change in the eligibility coefficients. I apply the procedure to the reduced-form event-study estimates, first averaging the effects in 2010–11 and then averaging the effects in 2014–18.

Appendix Table A8 and Figures A8 and A9 report the full sensitivity profiles. Section 5.4 reports the conventional averages, the robust intervals, and the values of $\bar{M}$ at which the confidence sets first include zero.

Together, these exercises distinguish three different restrictions on the counterfactual path: allowing a separate full-period linear trend, imposing exact continuation of the estimated pre-reform path, and bounding post-reform departures relative to the observed pre-reform changes. Section 5.4 uses them to assess the evidence around implementation separately from the evidence on longer-run persistence.

4.3.2 Spatial spillovers

The treatment and outcome are both defined over local markets rather than administrative village boundaries. Access is calculated from inhabited locations, and the segregation market contains recognised primary schools within one kilometre of the village polygon. A school added in response to an eligible village may therefore enter the feasible menu of residents in a nearby ineligible village. If such exposure increases access and affects sorting in comparison villages, the full-sample eligible–ineligible contrast will be attenuated.

I examine this possibility using a clean-comparison sample defined entirely from pre-reform geography. The restriction retains all baseline-eligible villages. Among baseline-ineligible villages, it retains only those for which none of the residents' one-kilometre school markets overlaps a 2008 target cell. The target cells are inhabited locations lying more than one kilometre from the nearest active public primary school in 2008, as defined in Appendix A.3. The overlap measure is constructed from resident cells rather than village-centroid distance. This screen retains 41.8 percent of the ineligible villages.

Appendix Table A6 reports the first-stage, reduced-form, and 2SLS estimates in this clean-comparison sample. Section 5.6 discusses how the resulting changes in the first stage and reduced form inform the possible importance of cross-boundary exposure. Because the clean screen also changes the composition of the comparison group, I retain the full sample as the preferred specification.

4.3.3 Exclusion restriction

To interpret β as the effect of multiple-school access, eligibility must affect caste allocation through the access change represented by X_{vt}. The institutional basis for the instrument is specific: eligibility is determined from the distance between inhabited locations and the pre-reform public-school stock, and the first-stage evidence shows a break in nearby public-school access at implementation. District-by-year fixed effects absorb district-wide implementation and other changes common to eligible and ineligible villages within the same district and year.

The exclusion condition is most naturally interpreted in relation to the full response of the local schooling market. An expansion in access may change enrolment, induce entry or adjustment by private schools, alter the allocation of pupils among public schools, or affect inputs at incumbent schools. These responses are channels through which a change in feasible choice can alter segregation. Including them as contemporaneous controls would condition on outcomes of the policy and remove part of the effect measured by equation (9). I therefore study enrolment, private-school responses, and incumbent-school inputs in Section 6. In particular, the specification that adds observable school inputs is presented there as part of the mechanism analysis rather than as a preferred control set.

The access results also clarify the scope of the 2SLS interpretation. Eligibility changes the distribution of nearby public-school menus, while X_{vt} summarises the policy-induced change at the at-least-two margin. The 2SLS estimate attributes the reduced-form change in segregation to this induced expansion in multiple-school access. The reduced form requires no such scaling and remains the direct estimate of how the eligibility rule changed caste allocation, conditional on the counterfactual-trends and spatial-interference assumptions. I therefore report the first stage, reduced form, and 2SLS estimate together throughout the paper.

4.3.4 Monotonicity

With heterogeneous causal responses, interpreting the continuous-treatment IV estimate also requires monotonicity in the instrumented treatment (Imbens and Angrist, 1994; Angrist and Imbens, 1995). In this setting, eligibility must weakly increase a village's potential share of residents with access to at least two nearby public primary schools. Equivalently, residents who would have had access to at least two schools if their village were ineligible should not lose that access because the village is eligible.

The condition concerns the cumulative at-least-two measure, not each exact menu category. For example, a movement from exactly two nearby schools to exactly three reduces the exact-two share but increases neither non-compliance nor defiance at the at-least-two margin. A violation would instead require eligibility to reduce the access level that would otherwise have been available, perhaps by displacing school provision that would have occurred in the absence of the rule. Potential access under the alternative eligibility status is not observed, so this comparison cannot be made village by village. The positive aggregate first stage, the timing of the access reversal, and the cumulative nature of the public-school stock are consistent with the monotonicity condition for the treatment used in the paper.

4.4 Interpretation of the estimates

The reduced-form coefficient τ is the intent-to-treat effect of eligibility on segregation in the established-market sample. It answers whether villages that fell short of the one-kilometre norm in 2008 experienced a different change in caste allocation after implementation than villages in the same district that already met the norm. Its causal interpretation rests on the counterfactual-trends and spatial-interference conditions discussed above, but not on the exclusion restriction used to scale the effect by multiple-school access.

The 2SLS coefficient β divides the reduced-form effect by the eligibility-induced change in X_{vt}. It is expressed per ten-percentage-point increase in the share of residents with at least two public primary schools within one kilometre. With a continuous treatment and heterogeneous effects, β is a weighted average of causal responses over the changes in access induced by eligibility (Angrist and Imbens, 1995). Because the outcome spans all recognised schools in the local market, this response can include reallocation among public schools and between public and private schools.

The estimand is specific to the design sample. It describes an increase in multiple-public-school access within markets that contained at least two recognised primary schools before the reform. It is therefore an effect of expanding feasible choice in established school markets, rather than an estimate of the effect of supplying a first school. The reduced form provides the policy effect of eligibility in those markets; the 2SLS estimate expresses that effect relative to the induced increase in access to at least two public schools. Section 5 reports both estimates.

5 Public-School Access and Caste Allocation

The results come in two parts. I first use the broad policy sample to estimate how baseline eligibility changed nearby public-school access and the composition of residents' feasible menus. I then turn to villages whose local markets already contained multiple primary schools before the reform and ask whether policy-induced growth in access to multiple public schools changed the allocation of SC and non-SC pupils across those schools. The separation keeps the formation of local school markets distinct from allocation within markets that already existed.

5.1 Public-School Access and Market Formation

The one-kilometre rule applies to habitations, not to villages. A village can hold a school at its centre and a hamlet at its edge lying well beyond a kilometre from it, so the rule can bind even where the village is not, in any ordinary sense, unserved. The places where this happens are large and spread out. Among the established-market villages in Table 1, eligible villages average 5.6 square kilometres against 1.8 for ineligible ones, and roughly twice the population. They also start with less provision: 4.2 nearby public primary schools against 6.0, and two-or-more access of 0.34 against 0.71. Eligibility therefore selects large villages with relatively thin provision.

Table 2 estimates equation (10) on the broad policy sample of 524,639 villages, 598 district clusters, and 4,769,014 village–year observations. Membership in this sample does not depend on whether segregation can be measured, so these estimates describe what the rule did to access

across rural India, not only in the markets studied below. Eligibility raised the share of residents with at least one public primary school within one kilometre by 6.92 percentage points against a mean of 0.846, and the share with at least two by 2.71 points against a mean of 0.488. The share with exactly two rose by 4.50 points against a mean of 0.236. The three F statistics are 178.9, 30.9, and 129.8.

The rule was written to deliver a first school, so the question that matters for this paper is whether it also changed access beyond that first school. Because the categories nest, the coefficients show where in the menu distribution the expansion landed. They imply a 6.92-percentage-point decline in the share of residents with no nearby public school, a 4.21-point increase in the share with exactly one, a 4.50-point increase in the share with exactly two, and a 1.79-point decline in the share with three or more. The expansion was therefore concentrated in one- and two-school menus. These are net changes in the share of residents occupying each state; they do not identify which residents moved between states, and the decline at three or more is a relative change rather than an absolute loss of schools.

For the rest of the paper, the relevant margin is therefore access to multiple nearby public schools. I next ask whether policy-induced growth on that margin changed how pupils were distributed across schools in markets where an across-school allocation margin already existed.

5.2 Caste Allocation Across Established School Markets

The estimates below use the established-market sample defined in Section 4.1: villages whose local market held at least two active recognised primary schools, public or private, in both 2006 and 2007. Membership is fixed on pre-reform structure, so these are intensive-margin estimates; they exclude villages in which the reform created a multi-school market for the first time. Contact outcomes remain defined in single-school markets, so Appendix D also examines the broader population, including markets without multiple schools at baseline; those estimates concern a different population rather than merely a robustness check on the established-market result. The endogenous variable is the share of residents with at least two nearby public primary schools. Outcomes are measured across all active recognised primary schools in the market, public and private, so the index registers movement among public schools as well as between the two sectors. Table 3 reports all three outcomes on one sample of 350,331 villages, 554 district clusters, and 3,185,709 village–year observations.

Eligibility raised the share of residents with at least two nearby public primary schools by 3.25 percentage points against a mean of 0.532, with a first-stage F statistic of 35.9. The corresponding effect is 2.71 percentage points in the broad policy sample; the difference reflects the distinct populations represented by the two estimates.

Expanding access to multiple public schools made the allocation of SC and non-SC pupils across the market's schools more uneven. Eligibility raised observed dissimilarity by 0.0054. The corresponding effect on dissimilarity expected under random allocation is -0.0002 and indistinguishable from zero, so dissimilarity beyond the random benchmark rose by 0.0057. Scaling by the access expansion, a ten-percentage-point increase in access to at least two nearby public schools raises observed dissimilarity by 0.0167, with an interval of $[0.0122, 0.0212]$, and beyond-random dissimilarity by 0.0174, with an interval of $[0.0125, 0.0223]$. The random benchmark

moves by −0.0007 and its interval spans zero.

The 2SLS estimate is local to the policy-induced increase in residents' access to at least two nearby public schools within established multi-school markets. It is not the effect of adding one school to every village or of creating a multi-school market for the first time. Rather, it captures the change in caste allocation associated with expanding multiple-public-school access where an across-school allocation margin already existed.

The random benchmark addresses a separate concern. It holds fixed the realised set of schools and their enrolments and reallocates caste at random across the school places they contain. Dividing a finite pupil population across more or smaller schools can mechanically raise measured dissimilarity even under random assignment; the benchmark asks whether that arithmetic accounts for the observed response. It does not: the estimated effect on random dissimilarity is close to zero. The exercise therefore captures the mechanical component of the index and nothing more. It is not a decomposition of the reform into entry and behaviour, and it does not identify the mechanism behind the remaining increase.

The magnitude is modest but economically meaningful. A ten-percentage-point increase in access to at least two nearby public primary schools raises observed dissimilarity by 0.0167. Relative to the 2008 mean among eligible villages of 0.391, this represents a 4.3 percent increase. The dissimilarity index has a direct reallocation interpretation: the estimate raises by 1.67 percentage points the share of SC pupils who would need to be reassigned across schools to give SC and non-SC pupils identical distributions. This is equivalent to roughly 17 SC pupils per 1,000. The corresponding beyond-random estimate of 0.0174 is 5.7 percent of its mean and 9.1 percent of its 2008 cross-village standard deviation. Although the measures and institutional settings differ, the estimate is close in relative magnitude to that of Monarrez et al. (2022), who find that charter-school expansion accounts for approximately 6 percent of mean segregation in US school districts.[16]

5.3 Cross-Caste Exposure Within Schools

The segregation estimates describe how pupils are distributed across schools. I next ask what this allocation means for the school environment experienced by a child. Expanding the set of nearby schools need not bring children from different caste groups into the same school. It may instead allow them to attend separate schools within the same local market. The relevant distinction is between access to schooling and exposure to the other group within school.

I measure this exposure in three ways: the number of out-group schoolmates, their share of primary enrolment, and whether the school contains any out-group pupils at all. The count describes how many pupils from the other group share the school; the share distinguishes composition from size; and exact zero identifies schools in which even co-presence is absent. Schools receive weights proportional to their enrolment of the focal group, so these outcomes describe the school of the average SC or non-SC pupil within each market.[17] Table 4 uses the same pre-

[16] School-level indices may not capture every adjustment in social exposure. Crema (2026), for example, finds that charter-school openings also increase segregation across classrooms within nearby traditional public schools. Crema nevertheless concludes that incorporating this additional margin is unlikely to overturn the literature's characterization of charter-related segregation effects as modest in magnitude.

[17] This follows the distinction between evenness and exposure in Massey and Denton (1988). Pupil weights

reform established-market restriction as the segregation analysis, maintaining the comparison within markets where an allocation margin already existed.[18]

In the pooled estimates, a ten-percentage-point increase in access to at least two nearby public primary schools reduces out-group schoolmates by 5.81 for SC pupils and 2.16 for non-SC pupils. That is roughly six fewer non-SC schoolmates for an SC pupil and two fewer SC schoolmates for a non-SC pupil, or 6.3 and 7.1 percent of their respective 2008 means of 91.80 and 30.33.[19] Fewer out-group schoolmates can, however, reflect smaller schools without any change in caste composition. The headcount estimates alone therefore cannot establish that pupils became less likely to share school with the other group.

The composition measures help distinguish these possibilities. For SC pupils, the non-SC share of their school falls by 0.91 percentage points, or 1.4 percent of its baseline mean. Their schools are not merely smaller in the pooled estimates: they are also more heavily SC. For non-SC pupils, the estimated decline in the SC share is 0.20 percentage points, with a standard error of 0.13, and is not statistically distinguishable from zero. Its confidence interval nevertheless includes a proportional decline as large as that estimated for SC pupils. The school-size and own-group counts are reported alongside these measures, but their sensitivity to specification prevents a stable numerical division of the exposure response into size and composition.

The clearest extensive-margin result is for non-SC pupils. Complete separation was already common before the reform: in 2008, the mean share attending schools with no SC pupils was 19.2 percent. Randomly allocating pupils across the same school places, holding school sizes and market caste totals fixed, would produce a mean of 2.7 percent. Complete separation was therefore roughly seven times as common as this finite-school benchmark would predict.[20]

A ten-percentage-point increase in multiple-school access raises the non-SC zero-contact share by 0.66 percentage points, or 3.4 percent of its baseline mean. At this margin, the school offers no opportunity for SC/non-SC interaction among its primary pupils. The increase beyond the random-allocation expectation is larger, at 0.89 percentage points: changes in school sizes and market caste totals alone do not generate the observed increase. The non-SC result holds under all five reported low-contact definitions, based on both shares and headcounts.[21] For SC pupils, the pooled zero-contact estimate is positive at 0.28 percentage points, from a lower baseline mean of 2.85 percent, but less stable across specifications. The evidence does not establish a symmetric increase in complete separation.

The distinction between these findings also matters for identification. Allowing an eligibility-

apply within markets; regressions and reported baseline means weight village–year observations equally. These are not national pupil-weighted totals. Exposure is measured across primary grades, not within a particular classroom.

[18] The contact sample also requires both groups in 2008 and complete group-specific outcomes. Its realised observations therefore differ from those in Table 3. Sample sizes and first stages are reported by group. Appendix Section D also reports a broader predetermined sample retaining single-school markets.

[19] The ten-percentage-point change is a common scaling of the IV estimates, not the access increase delivered by eligibility. Eligibility raised multiple-school access by approximately 3.3 percentage points in these samples; the corresponding reduced-form count effects are -1.91 and -0.71.

[20] This comparison describes the baseline allocation, not an effect of the reform. The random benchmark does not preserve residential proximity or household choices and is not a feasible school-siting counterfactual.

[21] Beyond-random effects range from 5.4 to 10.5 percent of their respective excess baseline means (Appendix Table A12). These are not percentage changes in the observed zero-contact rate. The largest estimate is not preferred; exact zero remains the main illustration.

specific linear trend substantially attenuates the reduced-form headcount effects, while leaving the SC out-group-share and non-SC observed zero-contact effects similar in magnitude. The adjustment also weakens the first stage, so larger trend-adjusted IV ratios should not be read as stronger evidence. Nor does stability under that adjustment resolve the pre-period differences in the SC share. Non-SC out-group counts and beyond-random zero contact have more reassuring pre-period diagnostics in established markets, although these tests cannot establish parallel counterfactual trends. The appendix reports the full diagnostics. I therefore treat the headcounts as an intuitive description of the pooled response, with greater sensitivity than the composition and complete-separation findings.

These outcomes give the segregation result a child-level interpretation, but not a direct welfare valuation. In Delhi schools, Rao (2019b) finds that exposure to poor classmates makes richer pupils more generous and less likely to discriminate against poor children. This illustrates why the presence of the other group can matter, beyond its numerical share. The form of interaction also matters: Lowe (2021b) finds that collaborative contact increases cross-caste friendships and reduces own-caste favouritism, while adversarial contact generally reduces cross-caste interaction.[22] The present data establish who shares a school, not whether pupils interact or how those interactions affect behaviour. The contribution is to show that access expansion can change the opportunities for cross-caste contact alongside school availability. Whether those changes translate into discrimination, friendships or a net welfare cost remains outside what these estimates identify.

5.4 Dynamics

Figure 2 reports the annual estimates from equation (13). Because eligibility is measured from the 2008 access environment, the figure uses 2008 as the omitted year and reports 2009 separately as the transition year between enactment and implementation. Panel (a) applies the specification to the two access margins in the broad policy sample; panel (b) applies it to the three segregation outcomes in the established-market sample.

The annual estimates place the change in caste allocation around enactment and implementation. For observed and beyond-random dissimilarity, the 2006 and 2007 coefficients are jointly insignificant, with p-values of 0.340 and 0.342. Their estimated pre-reform slopes are negative: -0.00079 per year for observed dissimilarity ($p = 0.076$) and -0.00072 for beyond-random dissimilarity ($p = 0.113$). Expected dissimilarity under random allocation shows little evidence of differential movement before the reform; its joint-test p-value is 0.892 and its slope is -0.00007 ($p = 0.704$). After 2008, observed and beyond-random dissimilarity become positive while the random-allocation benchmark remains near zero. The reduced-form effect on observed dissimilarity rises from 0.0038 in 2010 to 0.0053 in 2011 and remains around 0.006–0.007 thereafter.

The annual estimates also show that the response emerged quickly. By 2011, less than two years after implementation began, the eligibility coefficient for observed dissimilarity had

[22] Rao studies rich–poor integration in Delhi schools. Lowe randomly assigns Indian men to cricket teams and opponents, separating collaboration from competition. Corno et al. (2022b) exploit random roommate allocation at a South African university and find that interracial roommates reduce White students' negative stereotypes and increase interracial friendships. These designs concern specific forms of interpersonal contact, not a common response per out-group schoolmate.

reached 0.0053. This is about 60 percent of the absolute change in urban SC residential dissimilarity that Asher et al. (2026) document over the entire 2001–11 decade. The ten-percentage-point access estimate of 0.0167 is nearly twice the size of that decade-long change. Residential and school segregation are different outcomes, so I use this comparison only to provide a sense of time scale. Evidence from Indian schools points to similar persistence: Edara (2025b) finds that national caste-segregation indices changed little between 2007–08 and 2017–18. The policy-induced change was therefore relatively sharp against an otherwise slow-moving background.

The access series shows a reversal rather than a continuation of its earlier path. Before 2008, access to at least two nearby public schools was declining in eligible villages relative to ineligible villages, with an estimated differential slope of −0.0162 per year. Relative to 2008, the coefficient becomes positive in 2009 and reaches 0.0260 in 2010 and 0.0290 in 2011. The pre-reform movement is consistent with the institutional history in Section 2.1: habitation-level expansion was already under way through SSA, and the same one-kilometre threshold was already used to identify areas for new primary schools (Ministry of Human Resource Development, Government of India, 2004, p. 125). RTE therefore generated a break within an ongoing access expansion rather than initiating school construction from a static baseline. Extrapolating the earlier path forward strengthens rather than weakens the evidence of an implementation break: under exact continuation of the measured pre-reform path, the 95 percent intervals for the access effect exclude zero in 2009, 2010, and 2011.

I assess the sensitivity of the segregation result to this pre-reform movement in three ways. First, adding an eligible-specific linear trend reduces the reduced-form effect on observed dissimilarity from 0.0054 to 0.0033. The adjustment also reduces the first stage from 0.325 to 0.172, with an F statistic of 11.5, so the corresponding 2SLS estimate remains close to the baseline, at 0.0192 rather than 0.0167 (Appendix Table A7). Second, under exact continuation of the estimated pre-reform sorting path, the 95 percent intervals for observed and beyond-random dissimilarity exclude zero in 2009–11, while those for expected dissimilarity under random allocation include zero throughout (Appendix Figure A7). Third, under the relative-magnitude restrictions of Rambachan and Roth (2023), the confidence set for the average 2010–11 effect first includes zero at $\bar{M} = 0.95$ for observed dissimilarity and at $\bar{M} = 1.00$ for beyond-random dissimilarity. For the 2014–18 average, both confidence sets first include zero at $\bar{M} = 0.60$ (Appendix Table A8). The conventional early-period reduced-form estimate for observed dissimilarity is 0.0046. These sensitivity results apply to the stated period-specific reduced forms, not to the pooled 2SLS estimate of 0.0167. I therefore place greatest weight on the change around implementation and interpret the later coefficients as evidence of persistence subject to greater uncertainty about the counterfactual path.

5.5 Robustness

The Census-2011 controls. Because the village characteristics used in the baseline specification were measured after implementation began, Appendix Table A5 removes all Census-2011-by-year interactions while holding the sample fixed. The first stage changes from 0.325 to 0.323, the reduced-form effect on observed dissimilarity from 0.0054 to 0.0055, and the corresponding 2SLS estimate from 0.0167 to 0.0172. The results are therefore not generated by

conditioning on these characteristics.

The 2009 transition year. The Act was passed in August 2009 and took force on 1 April 2010, so 2009 belongs fully to neither period. Districts were mapping habitations and identifying children during that year, and some construction had begun. Table A1 therefore reports the pooled specification under all three conventions. Excluding 2009, as the main specification does, gives an observed-dissimilarity IV estimate of 0.0167. Retaining it as pre-reform gives 0.0155; treating it as the first post-reform year gives 0.0174. Beyond-random estimates run from 0.0162 to 0.0181 across the same columns. The first stage is weakest when 2009 is coded as post, at 0.2822 with an F of 33.1, which is the pattern one would expect if part of the year is not yet treated. The convention moves the estimate by about 12 percent of its own size and leaves its sign and significance unchanged.

The market-definition line. Requiring two schools in 2006 and 2007 fixes the established-market sample before the reform, but alternative definitions give similar results. Table A2 draws the line two other ways: requiring an observed 2008 count of at least two where that count exists, and requiring two schools in each of 2006, 2007, and 2008. The observed IV estimate moves from 0.0167 to 0.0162 and 0.0161, and the beyond-random estimate from 0.0174 to 0.0170. The random component sits at or near zero in every column. The result does not turn on where the market-definition line falls.

The choice of index. Dissimilarity is one summary of an uneven distribution, and it is insensitive to which group is concentrated where. Table A3 applies two alternatives to the same sample. A ten-percentage-point increase in multiple-school access raises Theil's entropy index by 0.0122 (s.e. 0.0018) and the SC isolation index by 0.0060 (s.e. 0.0019), against means of 0.205 and 0.374. Dissimilarity records unevenness across schools, Theil records it relative to the market's own composition, and isolation records the SC share in the school of the average SC pupil. All three rise, so SC pupils become both more unevenly distributed and more concentrated with one another.

5.6 Shared School Markets

Village boundaries do not enclose school markets, and the paper's own measurement concedes as much: a school counts toward a resident's menu whenever it lies within a kilometre, whatever administrative unit contains it. The same logic applies to controls. A school built to serve a target hamlet in one village can fall inside the one-kilometre radius of residents in the village next door. Those residents are classified as controls but may nevertheless be partially exposed to the reform. Such exposure would attenuate the eligible–ineligible contrast.

I construct a cleaner comparison group from the resident-centred overlap measure. The restriction keeps every baseline-eligible village. Among baseline-ineligible villages, it retains only controls for which none of the residents' one-kilometre school markets overlaps a 2008 target cell. The measure works from resident-cell markets rather than distances between village centroids, which matters because two centroids can be far apart while the inhabited edges of

their villages nearly touch. The screen is severe: it retains 41.8 percent of control villages and cuts the sample from 3,185,709 to 2,376,688 observations (Table A6).

Both components of the IV ratio rise, and by similar proportions. The first stage increases from 3.25 to 5.40 percentage points and the reduced form on observed dissimilarity from 0.0054 to 0.0085. The random-allocation reduced form remains at zero, while beyond-random dissimilarity rises from 0.0057 to 0.0086. The IV estimates barely move: 0.0158 against 0.0167 on observed dissimilarity, and 0.0159 against 0.0174 beyond random. This joint movement is consistent with partial exposure attenuating both the first-stage and reduced-form contrasts while leaving their ratio roughly intact. Because the restriction also changes the comparison group and drops most controls, I do not interpret the larger reduced form and first stage beyond this attenuation pattern. The narrower conclusion is sufficient: the main result does not depend on treating every baseline-ineligible village as untreated.

Taken together, the results establish two facts that motivate the mechanism analysis. The reform increased access to multiple nearby public schools, and SC and non-SC pupils subsequently became more unevenly distributed across the schools available to them. The random-allocation benchmark shows that this increase is not the mechanical consequence of dividing pupils across more or smaller schools. What these results do not yet establish is why the changing school network produces that reallocation. Section 6 asks whether the answer lies in geography: as provision moves closer to where children live, does existing residential geography translate more strongly into school composition?

6 How Local Access Maps Residential Geography into Schools

Section 5 shows that the access mandate raised SC–non-SC dissimilarity across primary schools in established local markets. A ten-percentage-point increase in multiple-public-school access raises dissimilarity by approximately 0.017. This section asks what explains that increase. The answer I document is that the mandate made public provision more local, and that a caste-neutral nearest-school rule, applied to fixed residential geography and the realised school network, reproduces most of the increase using differences in caste composition across Census villages alone and nearly all of it once within-village differentiation is disciplined by independent habitation data. Differential caste-specific preferences are therefore not necessary to reproduce the aggregate increase.[23] The benchmarks that establish this are geographic assignment exercises, not observed enrolment assignments. They ask how much of the causal segregation response a common distance rule can reproduce without introducing caste-specific preferences.

The RTE distance norm provides a natural place to begin. It required a public primary school within approximately one kilometre of every habitation. Meeting that norm increased the density of the public-school network and changed which school was closest to different

[23]Survey-weighted IHDS-II responses provide complementary descriptive evidence. Among government-primary pupils, 68.7 percent of SC households and 71.2 percent of non-SC households cite either proximity or the absence of another available school as the most important reason for school choice. The corresponding shares reporting an attended school in the survey's less-than-one-kilometre category are 88.1 and 89.9 percent. These similarities do not validate nearest-school assignment as a model of individual choice, but they show that the common-distance benchmark does not depend on an evident caste difference in the stated importance of geographic access.

settlements. This matters because caste composition varies across Census villages and across habitations within them. As provision becomes more local, schools draw from smaller residential areas and their composition more closely reflects differences across settlements. Before turning to this geography, I confirm that the response is not a change in who is counted.

6.1 Market composition and reporting

A change in who enters the measured school market is a plausible alternative to reallocation across schools. Around seven million children remained out of school in 2006–07, and the reported out-of-school rate was 8.17 percent among SC children, compared with 6.97 percent among OBC children (Planning Commission, Government of India, 2008, Chapter 1, p. 3). If segregation rose because the reform changed who is counted, the enrolment data should show it. New schools might have drawn mainly SC children who had previously been out of school or enrolled outside the measured local market, changing market SC share and hence the index even under random allocation across schools. Alternatively, treatment might have changed the set of schools that record caste composition, so that the market measured after the reform differs from the market measured before it. Both accounts predict changes in the measured population that are visible before any question about its allocation across schools arises.

Neither appears quantitatively important. Table 5 reports the reduced-form effects of eligibility on caste-valid enrolment. Total enrolment in the market rises by about 45 pupils, of whom 10.1 are SC and 34.9 non-SC, so SC pupils are 22.5 percent of marginal enrolment against 24.7 percent of the existing stock, and the aggregate SC share of the market falls by only 0.12 percentage points. Expected dissimilarity under random allocation, recomputed using the realised market population and school-size distribution in each year, changes by approximately -0.0002 and is statistically indistinguishable from zero (Column 2 of Table 3). Reporting is similarly stable: the number of schools reporting caste composition changes by 0.016 schools (s.e. 0.028), the reporting share by -0.09 percentage points (s.e. 0.08), and reporting exceeds 99.9 percent throughout (Appendix Table A4). The caste composition and reporting coverage of the measured population are broadly stable. What changes is how pupils are divided across schools.

6.2 The localization of public provision

Because the mandate directly expanded public provision, I first ask whether the public-school network became more local. By *more local*, I mean that the nearest public primary school became closer and more likely to lie within the resident's own Census village. Using the inhabited 100-metre GHSL cells introduced in the data section, I identify the closest active public primary school to every inhabited cell in 2008 and 2018 and calculate the distance to it. I then overlay both the cell and that school on Census-village boundaries and compare their village identifiers. This records whether the closest public school lies in the same Census village as the resident or in a neighbouring village. I aggregate the cell-level measures to the village-year level using cell population weights and estimate the main reduced-form specification on the resulting

outcomes.[24]

Eligibility brought public provision closer to where people live. Appendix E.1 reports the exercise. Eligibility reduces a village's population-weighted mean distance to the nearest public school by 64.6 metres (s.e. 5.4) from a baseline mean of 536 metres, a reduction of about 12 percent. It reduces by 3.8 percentage points (s.e. 0.34), from a baseline of 44.6 percent, the population share whose nearest public school lies in another Census village, a reduction of 8.5 percent; the own-village share rises by the same amount. Before the reform, the nearest public school lay across a village boundary for nearly half of residents, and the mandate moved roughly one in twelve of that group to a nearest school within their own village. A public school located in one village can be the closest school for residents of several neighbouring villages. The expansion reduced the population for whom the geographically closest public provision crossed a Census-village boundary, and therefore reduced the geographic area over which a common public school was the nearest provider.

This localization need not affect caste segregation. If the settlements previously pooled by a common nearby school had identical caste composition, separating their closest-school areas would leave simulated school composition unchanged. They do not have identical composition. Census villages differ in SC share, while the independent habitation records introduced below show substantial caste differentiation below the village level and indicate that higher-SC habitations tend to lie farther from the settlement core. As provision becomes more local, schools draw from smaller residential areas and their composition more closely reflects differences across settlements. The next exercise asks how much of the observed causal response can be reproduced by this change alone.

6.3 A coarse residential-geography benchmark

I answer this with a geographic assignment benchmark that combines four ingredients: the actual locations of population in the inhabited 100-metre cells; the observed SC share of each Census village; the realised locations and opening years of public primary schools; and one common rule that assigns every resident to the nearest active public school. Distance is the margin targeted by the policy, and applying the same rule to every resident keeps the benchmark caste-neutral. The rule describes geographic proximity, not school attendance. Rural India has no binding Census-village attendance zones, and parents may send a child to a school outside the village or bypass the nearest school. The benchmark does not assume that households attend the nearest school. It asks what the changing configuration of schools implies for composition under a common geographic rule and shows how much of the observed response that rule can reproduce, without attributing the remaining gap to any single mechanism.

SC and non-SC residents of the same cell are assigned by the same rule to the same school. Residential composition is held fixed at its Census values throughout, so nothing in the simulation allows households to relocate or to respond to schools. Over time, only the set and locations of active schools change as new schools open. If simulated market segregation moves

[24] Population weights are used to aggregate cells within each village-year. The regressions retain the village-year as the unit of observation and do not weight villages by their total population. Construction of the localization outcomes is described in Appendix E.1.

with eligibility, it does so because the expansion repartitions the same residential population across a different configuration of schools.

I begin with a deliberately restrictive version of the benchmark. Every inhabited cell inside a Census village is assigned that village's observed SC share. All cells in a village therefore have identical composition, and the benchmark imposes complete caste mixing within villages. The purpose of this simplification is to isolate differentiation across Census villages. Dividing residents of the same village among additional schools cannot change their simulated composition because every cell carries the same SC share. The benchmark can respond only when changes in the set and locations of active schools alter the pooling of populations from differently composed Census villages. If this version reproduces the response, it does so without using any information about where caste groups live within a village.

From the assigned populations, I reconstruct each school's SC share and each market's dissimilarity, then estimate the main reduced-form specification on the simulated index exactly as on the observed one. Both are estimated on the same first-stage-locked spatial sample, so comparing reduced forms is equivalent to comparing IV effects. On this sample, the observed reduced form is 0.0055 (s.e. 0.0008), slightly above the headline estimate of 0.0054 reported in Section 5 for the full established-market sample.

Column 2 of Table 7 reports the result. Differences in caste composition across Census villages, combined with the treatment-induced localization of the network, reproduce 85.5 percent of the causal segregation response, even though the benchmark imposes complete caste mixing within every village. The reduced form on simulated dissimilarity is 0.0047 (s.e. 0.0006) against the observed 0.0055, and the gap of 0.0008 is within its standard error of 0.0007 ($p \approx 0.22$). The coarse benchmark's point estimate is nonetheless 15 percent below the observed response. One source of this gap may be the within-village geography that the benchmark deliberately discards. If differently composed habitations occupy different parts of the settlement, a school opened to serve a previously distant habitation can be associated with a population whose caste composition differs from the Census-village average. Evaluating this possibility requires information on residential composition below the village level.

6.4 Within-village geography from habitation data

The Census does not report caste composition below the village level, but an independent nationwide administrative source records SC, ST, and general-category population for roughly 1.67 million named rural habitations.[25] The habitation records do not carry household coordinates, so I approximate habitation locations from geotagged non-school facilities that carry habitation names. The matching procedure, spatial checks, and coverage audit are reported in Appendix E.2; they use no school locations, treatment assignment, or school outcomes. Among habitations that receive reliable facility anchors, higher-SC habitations tend to lie farther from the settlement core. Within calibration villages, strict SC–non-SC dissimilarity across habitations averages 0.245: roughly one-quarter of one group would have to be reallocated across

[25]The source is the habitation directory maintained under the Jal Jeevan Mission for rural water supply. Its purpose is unrelated to schooling, and it was compiled without reference to school locations or enrolment. The directory is recent relative to the school panel. I use it as a proxy for persistent underlying settlement geography, not as a year-specific or pre-treatment measure.

habitations to equalise composition.

The facility anchors are not intended to recover the location of every household. They discipline a parsimonious residential rule in which the SC share of inhabited cells rises with distance from the village's population centre, while each village's total SC share is preserved exactly. A single parameter, λ, governs the strength of this within-village gradient, with $\lambda = 0$ corresponding to the coarse benchmark. I choose λ so that mean simulated residential dissimilarity matches the mean observed across habitations. The match occurs at $\lambda = 0.80$, with a district-bootstrap range of roughly 0.70 to 0.90 (Appendix E.2). Only after λ is fixed is the residential rule transported to spatially usable villages in the main estimation sample and combined with the realised set and locations of active schools. The habitation data discipline the residential input but play no role in selecting the final estimation sample or fitting the school-segregation response.

Adding within-village differentiation at the strength the habitation data imply raises the reproduced share from 85.5 to 98.5 percent. Column 3 of Table 7 reports the fine benchmark. At $\lambda = 0.80$, the reduced form on simulated dissimilarity is 0.0054 (s.e. 0.0006), and the gap of 0.00008 is well inside its standard error of 0.00067 ($p \approx 0.90$). Across the bootstrap range, the reproduced share runs from 95.4 percent at $\lambda = 0.70$ to 101.7 percent at $\lambda = 0.90$. With private schools in the assignment set, the $\lambda = 0.80$ benchmark reproduces 100.5 percent on its corresponding sample. The reproduced share therefore rises by 13 percentage points when the benchmark allows within-village caste differentiation at a strength selected from the habitation data. Because that parameter is chosen without using school locations or segregation outcomes, this refinement is not fitted to reproduce the treatment effect.

The calibrated rule does not provide the true location of caste groups in every village. Facility coordinates are proxies for habitation locations, and the radial relationship estimated in the anchored sample is transported to villages without matched habitations. The exercise instead shows that the result remains when within-village differentiation is disciplined by information that is independent of schools and school segregation.

6.5 Interpretation, alternative accounts, and limitations

The geographic benchmark reproduces most or nearly all of the aggregate response, but it does not rule out other changes that could affect measured dissimilarity or actual attendance. I examine three before interpreting the benchmark: the composition of newly opened schools, observable school inputs, and private-sector adjustment.

Incumbent schools. The most direct way to divide the same pupils more unequally is to add a school whose intake differs from the market average. An entrant that opens beside an SC hamlet and enrols mostly SC children raises market dissimilarity on its own, while every incumbent school retains its previous composition. Under this account, the market-level response is an artefact of the index absorbing one distinctive new unit, and the schools that existed before the reform should show no change in caste differentiation. The test is to recompute composition outcomes using only public schools operating before the reform and ask whether they, too, become more differentiated.

They do. Table 6 shows that the standard deviation of SC share across incumbent public schools rises by 0.0058 with eligibility, and the SC–non-SC attendance gap at incumbents classified as high-SC before the reform rises by 0.0090. These are repeated observations of school composition, not individual histories, and I do not observe which pupils enrol where in successive years; incumbent composition can change through successive cohorts as well as through transfers. The narrower inference is that the market response cannot be accounted for solely by adding an unusually composed entrant to the segregation index. Caste differentiation also increases within the pre-existing public-school network.

Observable school inputs. A possible explanation for the increase in segregation is that expanded access allows families to sort across schools with different levels of provision. Perceived quality is a salient consideration in Indian school choice. Among rural primary pupils attending private unaided schools, 51.6 percent cited the poor quality of a nearby government school as a reason for their choice, while 21.9 percent cited English-medium instruction and 14.8 percent cited better facilities (National Statistical Office, 2020, Table 53).[26] Although these responses concern private-school attendance, they establish school quality as an ex ante plausible margin of choice.

If SC and non-SC households respond differently to differences in school provision, caste segregation could increase even when families care primarily about school inputs rather than peer composition. The school census does not measure school effectiveness or every dimension of perceived quality, but it records several observable inputs: usable classrooms, boundary walls, drinking water, libraries, playgrounds, ramps, and books. I use these measures to ask whether the two caste groups were reallocated toward systematically different schools. Geographic proximity need not be the only attribute affecting allocation. The expansion may have changed the resources of new and existing schools, and SC and non-SC households may face different constraints in responding to those differences. I examine this possibility in two ways. First, I construct an input index for each incumbent public primary school using classrooms, boundary walls, drinking water, libraries, playgrounds, ramps, and books, and fix the index at its 2006–2008 mean. A ten-percentage-point increase in multiple-public-school access changes the non-SC-minus-SC gap in exposure to this predetermined index by −0.0006 (s.e. 0.0006), and SC pupils' relative attendance at schools in the bottom quartile of the pre-reform input distribution by −0.0004 (s.e. 0.0015) (Table 8). The expansion does not shift SC pupils toward observably worse incumbent schools relative to non-SC pupils.

Second, I examine contemporaneous inputs, which can themselves respond to entry. The index declines by 0.0106 for a ten-percentage-point increase in multiple-school access, relative to a mean of 0.618, a decline of about 1.7 percent, although its components move in different directions (Appendix Table A18). On the common sample, the segregation coefficient is 0.0164 without the contemporaneous index and 0.0161 when it is included, a change of less than 2 percent (Appendix Table A17). Because contemporaneous inputs are potentially outcomes of treatment, the controlled specification is diagnostic rather than a preferred causal estimate. Its stability indicates that measured changes in school resources do not account for the segregation

[26] Respondents could report multiple reasons, so these categories are not mutually exclusive and do not sum to 100 percent.

coefficient. The exercise cannot rule out differences in effectiveness, leadership, safety, language, or reputation that are not recorded in the school census, but measured inputs do not provide an alternative account of the main response.

Private-sector adjustment. The remaining margin is the boundary between public and private schooling. Because the headline segregation index covers the full local market, private entry or caste-differentiated public–private attendance could change measured allocation even if the public network were otherwise unchanged. Expanded public access produces neither substantial nearby private entry nor an increase in private enrolment. Conditional on a private school being present, a ten-percentage-point increase in public access reduces the private-attendance share by 0.0095 for SC pupils and 0.0186 for non-SC pupils. The non-SC–SC private-attendance gap therefore narrows by 0.0090 rather than widening, and the difference between private- and public-school SC shares changes by -0.0013 and is statistically indistinguishable from zero.

The main segregation response also persists in villages without a nearby private school at baseline, where private exit is least able to account for the result. The geographic benchmark points in the same direction: assigning residents to the nearest school of either type rather than the nearest public school changes the reproduced share by only a few percentage points. Private provision can interact with geography in particular markets, but selective movement into private schooling does not explain the aggregate response.

Interpretation. The mandate made the nearest public school closer and less likely to lie across a Census-village boundary. Applied to fixed residential geography, this change reproduces 85.5 percent of the observed response when villages are treated as internally mixed and 98.5 percent when the benchmark incorporates the degree of differentiation observed across habitations. Differential caste-specific preferences are therefore not necessary to reproduce the aggregate increase. This does not establish that households always attend the nearest school or that school choice is absent. Families may choose among old and new schools, public and private schools, and schools that differ in distance, inputs, language, safety, reputation, or peer networks. These choices may reinforce the geographic assignment or offset it. The data record school composition, not the origins and destinations of individual pupils.

School siting is part of the mechanism. Residential geography is treated as given, but the set and locations of schools are not natural features of the landscape. The distance norm, the decision to satisfy it through local school entry, and the realised locations that resulted are policy choices, and they remain inside the benchmark. The exercise is therefore not a decomposition between natural geography and household choice. It asks how far the changing configuration of schools, combined with a fixed residential pattern and a common geographic assignment rule, reproduces the observed treatment-induced change. The expansion changes how strongly residential geography is passed through into school composition.This distinction also separates the exercise from the minimum-distance decomposition in Monarrez (2023). Monarrez uses Census blocks, existing school sites, and attendance boundaries to decompose the cross-sectional level of US public-school assignment segregation. He finds that existing attendance boundaries make schools approximately 5 percent more integrated than assigning pupils to their nearest school, showing that institutional assignment can attenuate rather than simply transmit residential

sorting. Rural India has no binding attendance boundaries in this setting, and the object of interest here is not the level of segregation implied by a fixed configuration of schools. It is the causal change produced by an exogenous increase in network density.

6.6 Coverage and composition

The policy implication concerns where schools are placed, not whether access should expand. A distance standard sets a coverage requirement but does not, by itself, uniquely determine the configuration of school sites. Where alternative locations satisfy that requirement, their implications for cross-settlement pooling may differ. The geographic benchmark makes this a relevant planning question: how does the composition of the settlements served change when a school is placed at one feasible location rather than another? An appropriate comparison would also account for school capacity, available land, construction and staffing costs, and safe travel routes. Two configurations satisfying the same distance threshold need not provide the same travel convenience or level of provision.

The present estimates do not establish that a different configuration could have preserved the access gains while avoiding the increase in segregation. Nor do they value the resulting change in opportunities for cross-caste contact. They instead show why coverage alone is an incomplete description of the consequences of school placement. Evaluating whether feasible alternative sites could sustain more cross-settlement pooling, subject to comparable resource and access constraints, requires a separate counterfactual exercise. The implication is therefore to consider coverage and composition jointly when evaluating school locations, not to infer a net welfare loss or recommend fewer schools.

7 Conclusion

This paper asks whether bringing public schools closer to where children live changes who attends school with whom. India's Right to Education Act required a public primary school within one kilometre of every rural habitation, a rule written entirely in terms of distance and containing no assignment mechanism or explicit integration objective. Comparing villages that initially fell short of the rule with villages already covered, before and after implementation, I find that the expansion brought schools closer and made the allocation of Scheduled Caste and other pupils across nearby schools more uneven. In markets that already contained at least two schools, a ten-percentage-point increase in the share of residents with access to multiple public schools raises the dissimilarity index by about 0.0167. This is 4.3 percent of the 2008 pre-reform mean among eligible villages. In reallocation-equivalent terms, the estimate corresponds to roughly 17 additional SC pupils per thousand needing to attend a different school to restore the pre-reform degree of evenness. The magnitude is modest, comparable in proportional terms to the effect of charter expansion on segregation estimated by Monarrez et al. (2022), and it is not the arithmetic of dividing a fixed population across more schools: the expectation of the index under random allocation does not move.

The mechanism is spatial. When public schools are sparse, a single school may be the nearest feasible school for children in several settlements, and those children can be pooled in one

institution even when the settlements differ in caste composition. When the network becomes denser, schooling becomes more local. Each school draws from a smaller residential area, and its composition comes to reflect the composition of the settlements nearby. Holding residential patterns fixed and assigning every resident to the nearest school under one caste-neutral rule reproduces most of the segregation response from differences across villages alone, and nearly all of it once differentiation within villages is allowed at a level disciplined by independent settlement data. Families do choose among schools, and the benchmark is not a description of where any child actually enrols. What the exercise shows is that differences in caste-specific preferences are not needed to account for most of the aggregate increase. The density and arrangement of public infrastructure therefore shape how strongly residential segregation is translated into institutional segregation.

Seeing this required a different definition of the school market. Administrative approaches, by default, treat the Census village or a similar jurisdiction as the unit within which children attend school. But a school just across a village boundary can be closer than one inside it, and the same school can fall within the feasible local market of residents of several neighbouring villages. Because no national data record actual school catchments, I construct feasible school markets from the resident outward, combining a 100-metre population grid, geocoded school locations, and annual school records, and I allow those markets to cross administrative boundaries. Before the reform the nearest public school lay in another village for nearly half of rural residents. Cross-settlement pooling of this kind is a feature that sparse school networks can create and localization can weaken, and it is invisible in markets defined by administrative units. More broadly, administratively defined jurisdictions are convenient units for policy and data collection, but they need not describe the markets through which people actually use public services.

The evidence is about market-level allocation, not individual pupils, and it says nothing about what happened inside classrooms. Households may sort on peers, information, or unobserved school attributes correlated with composition, and the results do not rule this out; they show that residential geography and the school network can account for most of the aggregate response without requiring caste-specific differences in preferences. Nor do the results imply that the reform reduced welfare. The coverage gains were real, and the paper gives no reason to build fewer schools. What it shows is that an access standard defined by distance has a second consequence that evaluation typically omits. It changes who uses public institutions together. In a society where caste segregation in both residence and schooling has changed little over recent decades, and where primary school may be one of the few settings that create sustained contact across caste lines, universal coverage and social integration cannot be treated as objectives that advance together by default.

The implication extends beyond schools. Local public infrastructure does not only shorten the distance to a service; its spatial arrangement also helps determine which populations share the institutions that provide it. School construction therefore alters not only how far children travel but the spatial architecture through which social groups meet.

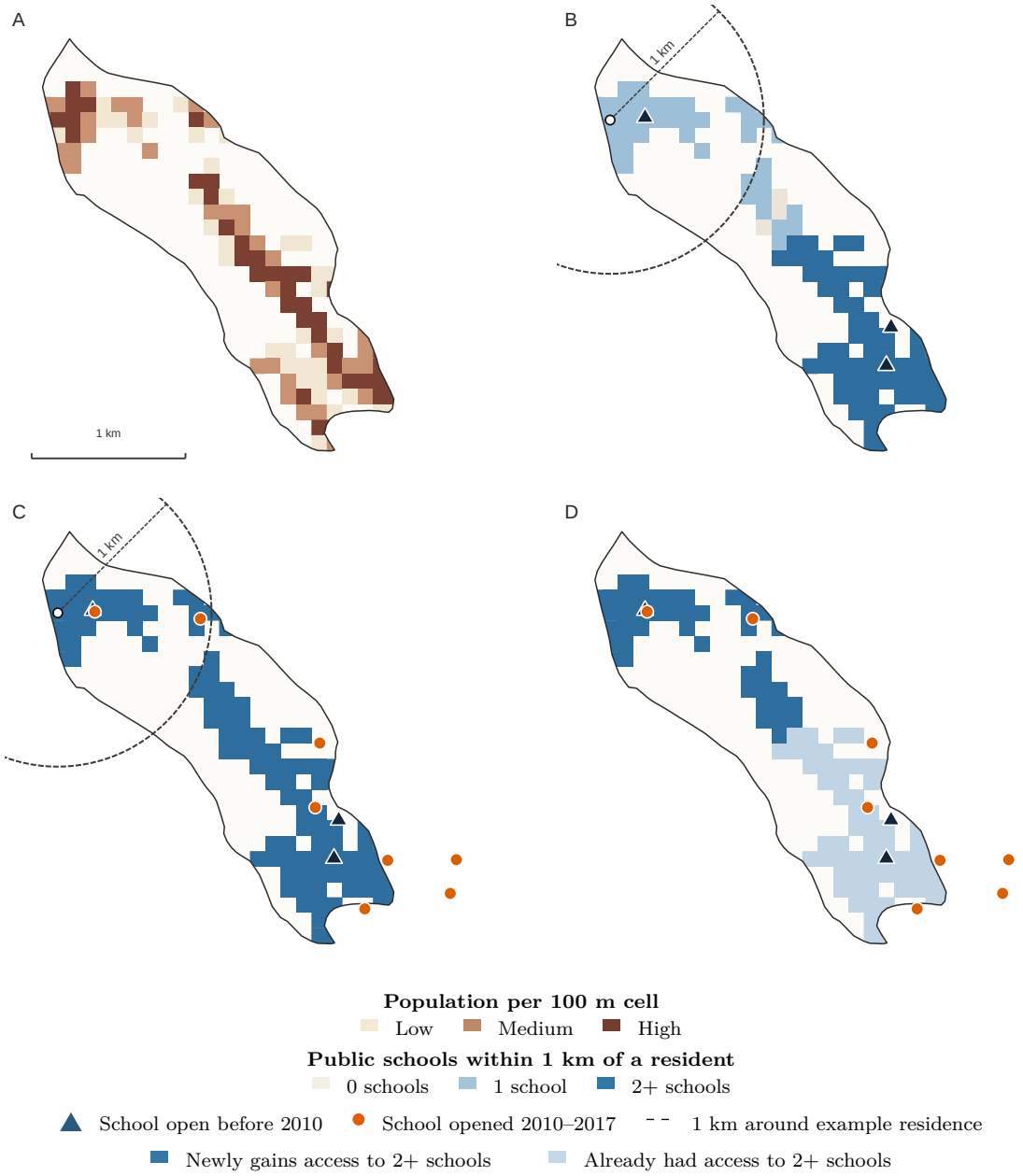
8 Tables & Figures

Table 1: Summary statistics

	Full sample	Established school markets		
	All	All	Ineligible	Eligible
	(1)	(2)	(3)	(4)
A. Access and school markets				
Share with at least one nearby public school	0.833	0.860	1.000	0.751
	(0.288)	(0.256)	(0.000)	(0.299)
Share with at least two nearby public schools	0.462	0.502	0.713	0.338
	(0.415)	(0.405)	(0.371)	(0.351)
Nearby public schools	4.596	4.963	6.008	4.153
	(3.538)	(3.449)	(3.973)	(2.717)
Nearby public and private schools	5.129	5.646	6.691	4.837
	(4.150)	(4.205)	(4.793)	(3.475)
Nearby public schools (count)	4.596	4.963	6.008	4.153
	(3.538)	(3.449)	(3.973)	(2.717)
Nearby private schools (count)	0.533	0.684	0.684	0.684
	(1.511)	(1.726)	(1.742)	(1.713)
B. Village characteristics				
Population	1463.868	1480.262	961.324	1882.349
	(2017.473)	(1972.992)	(1075.894)	(2375.215)
Area (sq. km)	4.036	3.947	1.768	5.635
	(6.510)	(5.847)	(2.046)	(7.135)
Scheduled Caste share	0.184	0.196	0.197	0.196
	(0.206)	(0.208)	(0.228)	(0.192)
Scheduled Tribe share	0.166	0.132	0.126	0.137
	(0.307)	(0.264)	(0.268)	(0.261)
All-weather road	0.792	0.798	0.754	0.833
	(0.406)	(0.401)	(0.431)	(0.373)
Bus access	0.448	0.462	0.346	0.553
	(0.497)	(0.499)	(0.476)	(0.497)
Mobile coverage	0.797	0.833	0.794	0.863
	(0.402)	(0.373)	(0.404)	(0.344)
Digital connectivity	0.827	0.859	0.821	0.888
	(0.378)	(0.348)	(0.384)	(0.315)
C. Baseline segregation				
Observed caste dissimilarity	0.365	0.411	0.438	0.391
	(0.232)	(0.209)	(0.206)	(0.208)
Expected dissimilarity under random allocation	0.098	0.113	0.119	0.109
	(0.096)	(0.100)	(0.108)	(0.093)
Dissimilarity beyond random allocation	0.268	0.298	0.319	0.282
	(0.202)	(0.191)	(0.184)	(0.194)
Villages	482,265	339,858	148,370	191,488

Notes: Column 1 describes the broad policy sample. This sample contains villages in the first-stage analysis universe with non-missing policy, identifier, and baseline-control variables; its construction does not condition on a segregation outcome or on the existence of multiple schools. Column 2 describes the common-outcome established-market sample, defined as villages with at least two active public or private primary schools within one kilometre in both 2006 and 2007. Columns 3 and 4 divide the established-market sample by baseline eligibility measured from the 2008 access environment. These columns are descriptive and are not presented as a randomisation-balance test. Entries are means, with standard deviations in parentheses. Variables are measured in 2008 unless they are time-invariant Census characteristics.

Figure 1: Resident-centred public-school access before and after the RTE expansion



Notes: The figure illustrates the resident-centred access measure for one Census village. The black line denotes the village boundary, while shaded cells are inhabited cells in the 2010 GHSL population grid. Panel A shows population density. Panel B classifies each inhabited cell by the number of public primary schools operating before 2010 that lie within 1 kilometre of its centroid. Panel C repeats the calculation after adding public schools that opened between 2010 and 2017. Navy triangles denote schools operating before 2010 and orange circles denote subsequently opened schools. Schools outside the village boundary are included whenever they lie within 1 kilometre of an inhabited cell. The dashed ring in Panels B and C illustrates a 1-kilometre radius around one residential location. The same resident-outward calculation is conducted separately for every inhabited cell; the ring is not an administrative school catchment. Panel D identifies inhabited cells that newly gain access to at least two public schools. The empirical measure is the population-weighted share of residents with at least two public schools within 1 kilometre. Population is held fixed at its 2010 distribution. The figure describes feasible school menus and does not identify pupil transfers.

Table 2: Baseline eligibility and access to nearby public schools

	Share of residents with public schools within one kilometre		
	At least one	At least two	Exactly two
	(1)	(2)	(3)
Eligibility × Post	0.0692***	0.0271***	0.0450***
	(0.0052)	(0.0049)	(0.0039)
Mean of outcome	0.846	0.488	0.236
F statistic	178.9	30.9	129.8
Villages	524,639	524,639	524,639
District clusters	598	598	598
Observations	4,769,014	4,769,014	4,769,014

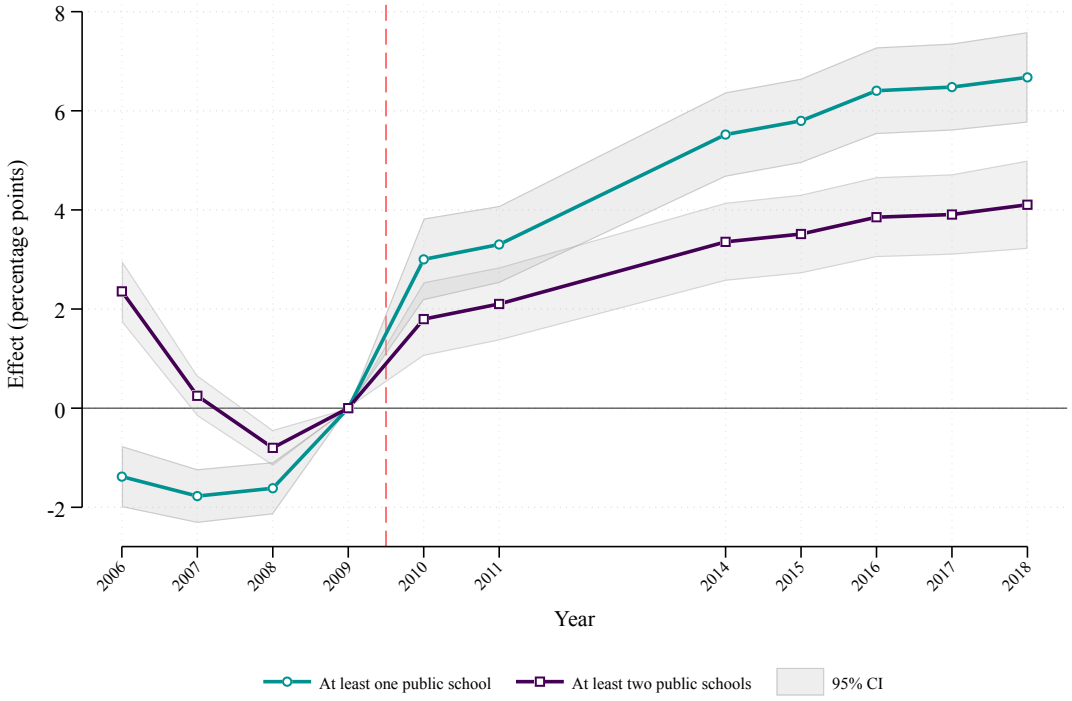
Notes: The table estimates how baseline eligibility changed residents' access to nearby public primary schools. The sample is the broad policy sample described in the notes to Table 1. Each column reports a separate regression of the outcome listed in the column on baseline eligibility interacted with an indicator for 2010 and later. The transition year 2009 is excluded. All specifications include village fixed effects, district-by-year fixed effects, and interactions between year indicators and baseline log population, log area, Scheduled Caste share, Scheduled Tribe share, all-weather-road access, bus access, mobile coverage, digital connectivity, a public distribution shop, and an ATM. Standard errors are clustered by district. Significance levels are denoted by $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table 3: Access to multiple public schools and caste dissimilarity

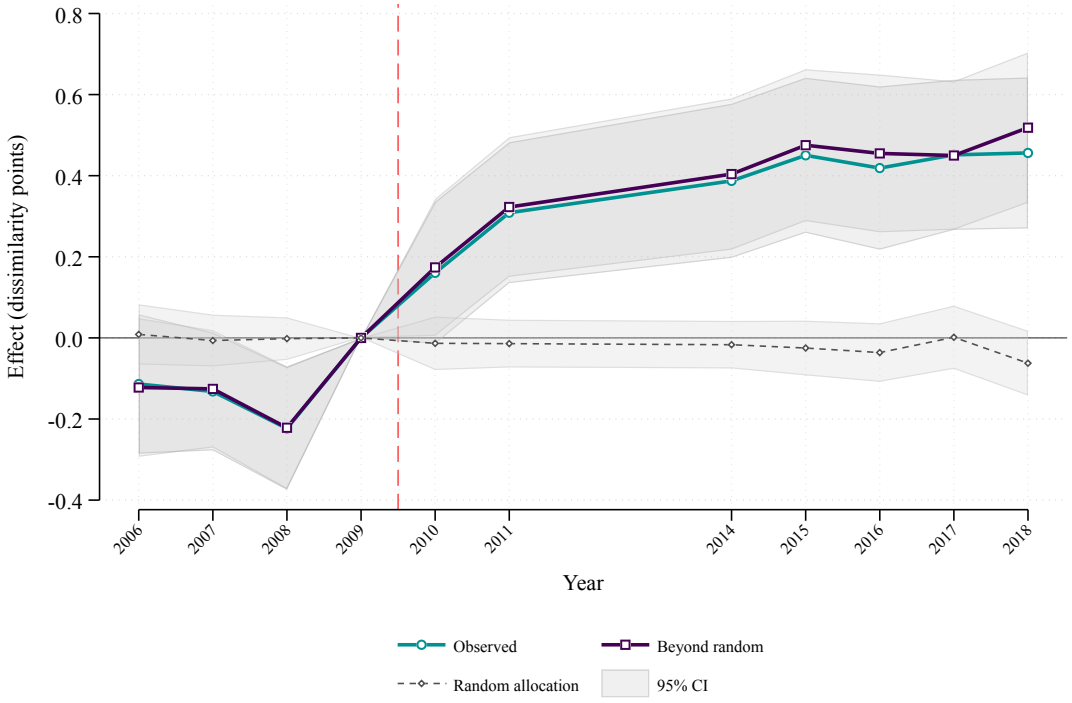
	Caste dissimilarity across local schools		
	Observed	Random allocation	Beyond random
	(1)	(2)	(3)
A. First stage (dependent variable: access to 2+ public schools, 10 p.p. units)			
Eligibility × Post	0.3253***	0.3253***	0.3253***
	(0.0543)	(0.0543)	(0.0543)
Effect in percentage points	3.25	3.25	3.25
Mean of outcome	0.532	0.532	0.532
First-stage F statistic	35.9	35.9	35.9
Observations	3,185,709	3,185,709	3,185,709
B. Reduced form			
Eligibility × Post	0.0054***	-0.0002	0.0057***
	(0.0008)	(0.0003)	(0.0008)
Mean of outcome	0.431	0.126	0.304
Observations	3,185,709	3,185,709	3,185,709
C. Two-stage least squares			
Access to 2+ public schools (10 p.p.)	0.0167***	-0.0007	0.0174***
	(0.0023)	(0.0009)	(0.0025)
	[0.0122, 0.0212]	[-0.0025, 0.0011]	[0.0125, 0.0223]
Mean of outcome	0.431	0.126	0.304
Villages	350,331	350,331	350,331
District clusters	554	554	554
Observations	3,185,709	3,185,709	3,185,709

Notes: The sample consists of villages with at least two active public or private primary schools within one kilometre in both 2006 and 2007. The 2008 school count does not determine sample membership. All three outcome columns use identical observations. Column 1 reports observed SC–non-SC dissimilarity across the full local public–private primary-school market. Column 2 reports expected dissimilarity from 100 random allocations that hold fixed the realised schools, school sizes, and market-level caste totals. Column 3 is observed dissimilarity minus its random-allocation expectation. Panel A reports the first stage for the share of residents with at least two public primary schools within one kilometre. Panel B reports reduced-form effects of baseline eligibility. Panel C instruments multiple-public-school access with baseline eligibility interacted with the post-reform period; access is scaled so that the coefficient is the effect of a 10 percentage-point increase. The transition year 2009 is excluded. All specifications include village fixed effects, district-by-year fixed effects, and the baseline-characteristic-by-year controls listed in the notes to Table 2. Standard errors are clustered by district. Parentheses contain standard errors; square brackets contain 95 percent confidence intervals. Significance levels are denoted by $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Figure 2: Dynamic effects of baseline eligibility on school access and caste segregation



(a) Access to nearby public schools



(b) Caste dissimilarity

Notes: The figure reports coefficients from interactions between baseline eligibility and year indicators. The omitted year is 2009, the enactment-to-implementation transition year. The access panel uses the broad policy sample and a common sample for access to at least one and at least two public primary schools within one kilometre. The segregation panel uses the established-market sample of villages with at least two public or private primary schools in both 2006 and 2007, and identical observations for observed dissimilarity, expected dissimilarity under random allocation, and their difference. All specifications include village fixed effects, district-by-year fixed effects, and baseline village characteristics interacted with year. Standard errors are clustered by district; shaded areas show 95 percent confidence intervals.

Table 4: Multiple-school access and cross-caste exposure

	Out-group schoolmates (1)	Out-group share (pp) (2)	Own-group schoolmates (3)	School size (4)	Zero out-group schoolmates (pp) (5)	Zero contact, excess (pp) (6)
Panel A. SC pupils						
Eligibility (RF)	−1.915***	−0.301***	0.254*	−1.661***	0.092*	0.095**
	(0.249)	(0.072)	(0.154)	(0.337)	(0.049)	(0.048)
Access, 10 pp (2SLS)	−5.807***	−0.914***	0.770	−5.037***	0.278**	0.288**
	(1.012)	(0.191)	(0.478)	(1.167)	(0.136)	(0.131)
2008 mean	91.80	63.57	45.13	137.93	2.85	2.83
IV / 2008 mean (%)	−6.3	−1.4	1.7	−3.7	9.8	10.2
Random zero-contact mean	—	—	—	—	0.017	—
First stage: 0.3297 (0.0553); clustered first-stage F: 35.50. 3,069,602 village–years; 335,709 villages; 541 district clusters.						
Panel B. Non-SC pupils						
Eligibility (RF)	−0.710***	−0.065	−0.638**	−1.348***	0.217**	0.293***
	(0.112)	(0.045)	(0.282)	(0.331)	(0.085)	(0.081)
Access, 10 pp (2SLS)	−2.157***	−0.197	−1.940**	−4.097***	0.658***	0.892***
	(0.414)	(0.127)	(0.900)	(1.127)	(0.215)	(0.203)
2008 mean	30.33	20.79	108.66	140.00	19.24	16.58
IV / 2008 mean (%)	−7.1	−0.9	−1.8	−2.9	3.4	5.4
Random zero-contact mean	—	—	—	—	2.667	—
First stage: 0.3290 (0.0552); clustered first-stage F: 35.49. 3,080,338 village–years; 335,824 villages; 542 district clusters.						

Notes: The sample uses the baseline established-market restriction, requires both caste groups in 2008, and retains complete outcomes and positive focal-group enrolment. All columns within a panel use a common sample. School outcomes are weighted by focal-group enrolment within markets; village–year observations are equally weighted. Out-group shares use total primary enrolment as the denominator. Own-group schoolmate counts exclude the focal pupil; school size includes that pupil. Excess zero contact subtracts the exact hypergeometric expectation conditional on market caste totals and school sizes. The models use village and district–year fixed effects and the baseline controls. Years are 2006–2018, excluding 2009, 2012 and 2013. Parentheses report district-clustered standard errors; stars use conventional, not weak-instrument-robust, inference: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The first-stage F is the squared clustered first-stage t statistic. Pre-period behaviour and trend sensitivity differ by outcome. In particular, school-size and own-group coefficients are specification-sensitive, and the SC zero-contact effect is not robust; see Appendix Section D. The relative-change row is 100 times each pooled IV coefficient divided by its own unrounded 2008 mean; it does not imply equal robustness across outcomes. The random zero-contact baseline is column (5) minus column (6), using unrounded means on the same sample; it is expressed in percentage points.

Table 5: Market enrolment and caste composition

	Total pupils (1)	SC pupils (2)	Non-SC pupils (3)	SC share (4)
A. First stage				
Eligibility × Post		0.3253***		
		(0.0543)		
First-stage F statistic		35.9		
B. Reduced form				
Eligibility × Post	45.0191***	10.1206***	34.8984***	-0.1163**
	(3.9259)	(1.0732)	(3.1113)	(0.0467)
C. Two-stage least squares				
Access to 2+ public schools (10 p.p.)	138.3923***	31.1117***	107.2806***	-0.3575**
	(22.7824)	(5.2116)	(18.0223)	(0.1576)
Mean of outcome	655.619	158.393	497.226	24.701
Villages		350,331		
District clusters		554		
Observations		3,185,709		

Notes: The table reports effects on enrolment and caste composition in the established-market sample, defined as villages with at least two active public or private primary schools within one kilometre in both 2006 and 2007. Columns (1)–(3) report total, SC, and non-SC primary enrolment across all active public and private schools in the local market. Column (4) reports the SC share of market enrolment, in percentage points. Panel A reports the first stage for the share of residents with access to at least two public primary schools within one kilometre. Panel B reports reduced-form effects of baseline eligibility interacted with the post-reform period. Panel C instruments multiple-public-school access with the same interaction; coefficients represent the effect of a 10-percentage-point increase in access. The transition year 2009 is excluded. All specifications include village fixed effects, district-by-year fixed effects, and baseline village characteristics interacted with year. Standard errors are clustered by district and reported in parentheses.

Table 6: Caste differentiation within incumbent public schools

	SD of SC share (1)	SC HHI (2)	High-SC gap (3)	Non-SC exposure (4)
A. First stage				
Eligibility × Post			0.3389***	
			(0.0555)	
First-stage F statistic			37.3	
B. Reduced form				
Eligibility × Post	0.0020***	-0.0144***	0.0031***	-0.0001
	(0.0005)	(0.0016)	(0.0007)	(0.0005)
C. Two-stage least squares				
Access to 2+ public schools (10 p.p.)	0.0058***	-0.0426***	0.0090***	-0.0003
	(0.0010)	(0.0039)	(0.0021)	(0.0013)
Mean of outcome	0.146	0.491	0.348	0.215
Villages			347,526	
District clusters			550	
Observations			3,063,581	

Notes: The outcomes are constructed using incumbent public primary schools, schools that were already operating before the reform. The sample is the subset of the established-market sample for which all four incumbent-school outcomes are available. Column (1) reports the enrolment-weighted standard deviation of SC enrolment shares across incumbent schools. Column (2) reports the Herfindahl–Hirschman index of the concentration of SC enrolment across incumbent school units. The HHI is a measure of the concentration of SC pupils across schools and should not be interpreted as a directional SC–non-SC segregation index. Column (3) reports the SC–non-SC difference in attendance at high-SC incumbent schools. Column (4) reports the SC share of peers experienced by the average non-SC pupil in incumbent schools. Panel A reports the first stage for the share of residents with access to at least two public primary schools within one kilometre, expressed in 10-percentage-point units. Panel B reports reduced-form effects of statutory eligibility interacted with the post-reform period. Panel C instruments multiple-public-school access with the same interaction; coefficients represent the effect of a 10-percentage-point increase in access. All four columns use identical observations. The transition year 2009 is excluded. All specifications include village fixed effects, district-by-year fixed effects, and baseline village characteristics interacted with year. Standard errors are clustered by district and reported in parentheses. These incumbent-school outcomes provide supporting evidence on changes within the pre-existing public-school network; they do not formally decompose the full-market segregation effect. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 7: Residential geography and the segregation response

	Observed school segregation	Geographic-assignment benchmark Nearest active public school	
	(1)	Coarse ($\lambda = 0$)	Fine, JJM-calibrated ($\lambda = 0.80$)
Eligibility × Post	0.00550***	0.00470***	0.00541***
	(0.00082)	(0.00055)	(0.00060)
Share of observed response	100.0%	85.5%	98.5%
Residual: observed minus benchmark	–	0.00080	0.00008
		(0.00065)	(0.00067)
First-stage F statistic	36.1	36.1	36.1
Villages	347,829	347,829	347,829
District clusters	540	540	540
Observations	3,169,565	3,169,565	3,169,565

Notes: Column 1 reports the reduced-form effect of baseline eligibility on observed SC–non-SC dissimilarity across schools. Columns 2 and 3 report reduced-form effects on dissimilarity generated by geographic-assignment benchmarks that assign every inhabited 100-metre GHSL cell to its nearest active public primary school. In Column 2, every cell inherits its Census village's SC share, imposing complete within-village mixing. Column 3 allows SC share to vary within villages according to the school-blind radial rule calibrated to match mean dissimilarity in the Jal Jeevan Mission habitation data, with the headline parameter fixed at $\lambda = 0.80$. The residual is the observed outcome minus the corresponding benchmark outcome, estimated directly in the same regression specification; its standard error is reported beneath it. The share of the observed response is the ratio of benchmark to observed reduced-form coefficients and is not an R^2 or a causal mediation share. All columns use the same first-stage-locked public-school spatial sample and include village fixed effects, district-by-year fixed effects, and year interactions with the baseline controls used in the headline specification. The transition year 2009 is excluded. Standard errors are clustered by district. Significance levels are denoted by $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.10$.

Table 8: Caste-specific exposure to fixed pre-reform observable school inputs

	SC exposure (1)	Non-SC exposure (2)	Non-SC − SC gap (3)	SC excess in low-input schools (4)
Eligibility × Post first stage	0.3221***	0.3221***	0.3221***	0.3221***
	(0.0547)	(0.0547)	(0.0547)	(0.0547)
First-stage F statistic	34.7	34.7	34.7	34.7
Eligibility × Post reduced form	0.0004**	0.0002	-0.0002	-0.0001
	(0.0002)	(0.0001)	(0.0002)	(0.0005)
2SLS: access to 2+ schools (10 p.p.)	0.0012**	0.0006	-0.0006	-0.0004
	(0.0006)	(0.0004)	(0.0006)	(0.0015)
Mean of outcome	0.526	0.522	-0.003	-0.005
Villages	342,625	342,625	342,625	342,625
District clusters	549	549	549	549
Observations	3,076,931	3,076,931	3,076,931	3,076,931

Notes: The unit of observation is a village–year in the predetermined established-market sample; 2009 is excluded. The outcomes are constructed using incumbent public primary schools only. Each school's observable-input index is fixed at its 2006–2008 average and is retained when the school has at least two usable pre-reform observations. Current grade 1–5 SC and non-SC enrolment provide the weights used to construct caste-specific exposure. Columns (1) and (2) report the enrolment-weighted input index experienced by SC and non-SC pupils, respectively. Column (3) is non-SC exposure minus SC exposure, so a positive coefficient indicates relative non-SC advantage. Column (4) is the SC share minus the non-SC share attending schools in the bottom quartile of the fixed pre-reform input index, so a positive coefficient indicates relative SC concentration in low-input schools. The reduced-form row reports the coefficient on baseline eligibility interacted with the post-reform period. The 2SLS row instruments the share of residents with access to at least two public primary schools within one kilometre using the same interaction; coefficients are reported per 10-percentage-point increase in access. All specifications use a common sample and include village fixed effects, district-by-year fixed effects, and predetermined village characteristics interacted with year. Standard errors are clustered by district and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

References

Adukia, A., Asher, S., and Novosad, P. (2020). Educational investment responses to economic opportunity: Evidence from indian road construction. *American Economic Journal: Applied Economics*, 12(1):348–376.

Afridi, F. (2011). The impact of school meals on school participation: Evidence from rural india. *Journal of Development Studies*, 47(11):1636–1656.

Allen, R., Burgess, S., Davidson, R., and Windmeijer, F. (2015). More reliable inference for the dissimilarity index of segregation. *Econometrics Journal*, 18(1):40–66.

Andrabi, T., Bau, N., Das, J., Karachiwalla, N., and Khwaja, A. I. (2024). Crowding in private quality: The equilibrium effects of public spending in education. *Quarterly Journal of Economics*, 139(4):2525–2577.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

ASER Centre (2005). Annual status of education report (rural) – various years. New Delhi: ASER Centre. and subsequent issues.

Asher, S., Jha, K., Novosad, P., Adukia, A., and Tan, B. (2026). Residential segregation and unequal access to local public services in india: Evidence from 1.5m neighborhoods. NBER Working Paper 34818, National Bureau of Economic Research. Author order was randomized.

Bailwal, N. and Paul, S. (2021). Caste discrimination in provision of public schools in rural india. *The Journal of Development Studies*, 57(11):1830–1851.

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.

Banerjee, A. and Munshi, K. (2004). How efficiently is capital allocated? evidence from the knitted garment industry in tirupur. *Review of Economic Studies*, 71(1):19–42.

Bassi, M., Besbas, B., Dinarte-Diaz, L., Ravindran, S., and Reynoso, A. (2026). From access to achievement: The primary school-age impacts of an at-scale preschool construction program in highly deprived communities. *American Economic Journal: Applied Economics*. Forthcoming.

Bau, N. (2022). Estimating an equilibrium model of horizontal competition in education. *Journal of Political Economy*, 130(7):1717–1764.

Böhlmark, A., Holmlund, H., and Lindahl, M. (2016). Parental choice, neighbourhood segregation or cream skimming? an analysis of school segregation after a generalized choice reform. *Journal of Population Economics*, 29(4):1155–1190.

Burde, D. and Linden, L. L. (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*, 5(3):27–40.

Carrington, W. J. and Troske, K. R. (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, 15(4):402–409.

Chaudhury, N. and Kingdon, G. G. (2014). School choice in india: A mixed-methods study on private schooling in rural areas. *Prospects*, 44(3):267–282.

Choudhury, P. and Bhattacharya, S. (2020). School choice and inequality in india: Patterns of private school enrolment. *International Journal of Educational Development*, 75:102170.

Corno, L., La Ferrara, E., and Burns, J. (2022a). Interaction, stereotypes, and performance: Evidence from south africa. *American Economic Review*, 112(12):3848–3875.

Corno, L., La Ferrara, E., and Burns, J. (2022b). Interaction, stereotypes, and performance: Evidence from South Africa. *American Economic Review*, 112(12):3848–3875.

Cortese, C. F., Falk, R. F., and Cohen, J. K. (1976). Further considerations on the methodological analysis of segregation indices. *American Sociological Review*, 41(4):630–637.

Crema, A. (2026). The segregative effects of charter schools. *American Economic Journal: Economic Policy*, 18(2):242–274.

Department of School Education and Literacy (2010). Web based school gis application. Ministry of Education, Government of India. School GIS portal for viewing geographic coordinates of schools.

Department of School Education and Literacy (2023). Unified district information system for education plus (udise+): Reference booklet. Ministry of Education, Government of India.

Desai, S. and Dubey, A. (2012). Caste in 21st century india: Competing narratives. *Economic and Political Weekly*, 47(11):40–49.

Desai, S., Vanneman, R., and National Council of Applied Economic Research (2018). India human development survey-ii (IHDS-II), 2011–12. Inter-university Consortium for Political and Social Research [distributor].

Deshpande, S. (2011). *Contemporary India: A Sociological View*. Penguin.

d'Haultfœuille, X. and Rathelot, R. (2017). Measuring segregation on small units: A partial identification analysis. *Quantitative Economics*, 8(1):39–73.

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

Dumont, L. (1980). *Homo Hierarchicus: The Caste System and Its Implications*. University of Chicago Press.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Edara, R. (2025a). Extent, patterns, and geographic scale of caste-based segregation in indian schools. *International Journal of Educational Development*, 117:103347.

Edara, R. (2025b). Extent, patterns, and geographic scale of caste-based segregation in indian schools. *International Journal of Educational Development*, 117:103347.

Edara, R. (2025c). Right to education (RTE) act's influence on caste-based enrollment gaps and segregation in india. EdWorkingPaper 25-1310, Annenberg Institute at Brown University.

Ferreyra, M. M. (2007). Estimating the effects of private school vouchers in multidistrict economies. *American Economic Review*, 97(3):789–817.

Government of India (2009). The right of children to free and compulsory education act, 2009. Act No. 35 of 2009, Gazette of India.

Gupta, D. (2000). *Interrogating Caste: Understanding Hierarchy and Difference in Indian Society*. Penguin.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jacoby, H. G. and Mansuri, G. (2015). Crossing boundaries: How social hierarchy impedes economic mobility. *Journal of Economic Behavior & Organization*, 117:135–154.

Joshi, P. (2024). Understanding parental perceptions and their involvement in children's learning through the lenses of accountability: Evidence from rural india. UKFIET blog.

Kaul, V., Bhattacharjea, S., Chaudhary, A. B., Ramanujan, P., Banerji, M., and Nanda, M. (2017). The india early childhood education impact study. Technical report, UNICEF, New Delhi.

Kazianga, H., Levy, D., Linden, L. L., and Sloan, M. (2013). The effects of "girl-friendly" schools: Evidence from the BRIGHT school construction program in burkina faso. *American Economic Journal: Applied Economics*, 5(3):41–62.

Kumar, A. (2023). Street safety: Fuel to girls' secondary schooling in india. Unpublished manuscript.

Lowe, M. (2021a). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–1844.

Lowe, M. (2021b). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–1844.

Lucas, A. M. and Mbiti, I. M. (2012). Access, sorting, and achievement: The short-run effects of free primary education in kenya. *American Economic Journal: Applied Economics*, 4(4):226–253.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, 67(2):281–315.

Ministry of Human Resource Development, Government of India (2004). *Sarva Shiksha Abhiyan: A Programme for Universal Elementary Education—Manual for Planning and Appraisal*. Department of Elementary Education and Literacy.

Ministry of Human Resource Development, Government of India (2010). The right of children to free and compulsory education rules, 2010. Model Rules under the Right of Children to Free and Compulsory Education Act, 2009.

Ministry of Human Resource Development, Government of India (2011). Guidelines under section 35(1) of the right of children to free and compulsory education act, 2009 regarding free and compulsory education in a neighbourhood school. Department of School Education and Literacy. Issued 25 July 2011.

Monarrez, T., Kisida, B., and Chingos, M. (2022). The effect of charter schools on school segregation. *American Economic Journal: Economic Policy*, 14(1):301–340.

Monarrez, T. E. (2023). School attendance boundaries and the segregation of public schools in the united states. *American Economic Journal: Applied Economics*, 15(3):210–237.

Mousumi, M. A. (2022). The public–private school decision in delhi, india. *London Review of Education*, 20(1):1–16.

Mukhopadhyay, A. and Sahoo, S. (2016). Does access to secondary education affect primary schooling? evidence from india. *Economics of Education Review*, 54:124–142.

Munshi, K. (2019a). Caste and the indian economy. *Journal of Economic Literature*, 57(4):781–834.

Munshi, K. (2019b). Caste and the indian economy. *Journal of Economic Literature*, 57(4):781–834.

Munshi, K. and Rosenzweig, M. (2006a). 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. (2006b). 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. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98.

Muralidharan, K. and Kremer, M. (2006). Public and private schools in rural india. In Patrinos, H. A. and Sosale, S., editors, *School Choice International: Exploring Public-Private Partnerships*, pages 91–110. World Bank.

Muralidharan, K. and Sundararaman, V. (2015a). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *Quarterly Journal of Economics*, 130(3):1011–1066.

Muralidharan, K. and Sundararaman, V. (2015b). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *Quarterly Journal of Economics*, 130(3):1011–1066.

National Sample Survey Organisation (2003). Report on village facilities, july–december 2002. NSS Report 487 (58/3.1/1), Government of India, New Delhi.

National Statistical Office (2020). Household social consumption on education in india. NSS Report 585, Ministry of Statistics and Programme Implementation, Government of India. NSS 75th Round, July 2017–June 2018.

Neilson, C., Dinerstein, M., and Otero, S. (2020). The equilibrium effects of public provision in education markets: Evidence from a public school expansion policy. Working Paper 645, Princeton University, Industrial Relations Section.

Office of the Registrar General and Census Commissioner, India (2011). Census of india 2011: Metadata and brief overview. Government of India. Administrative unit definitions and coding for villages and towns.

Planning Commission, Government of India (2008). *Eleventh Five Year Plan (2007–2012), Volume II: Social Sector*. Oxford University Press, New Delhi.

Programme Evaluation Organisation, Planning Commission, Government of India (1965). Problems of extension of primary education in rural areas. Technical report, Planning Commission, Government of India, New Delhi.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5):2555–2591.

Rao, G. (2019a). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3):774–809.

Rao, G. (2019b). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3):774–809.

Rathelot, R. (2012). Measuring segregation when units are small: A parametric approach. *Journal of Business & Economic Statistics*, 30(4):546–553.

Romero, M. and Singh, A. (2022). The incidence of affirmative action: Evidence from quotas in private schools in india. RISE Working Paper 22/088, Research on Improving Systems of Education.

Schiavina, M., Freire, S., MacManus, K., et al. (2022). Global human settlement layer (ghsl) data package 2022. Technical Report JRC126219, Joint Research Centre, European Commission. Includes 100 m resolution global population grids.

Shah, M. and Steinberg, B. (2019). The right to education act: Trends in enrollment, test scores, and school quality. *AEA Papers and Proceedings*, 109:232–238.

Singh, R. (2015). Determinants of successful completion of secondary school in india. Technical report, Young Lives Working Paper 142, Oxford.

Thorat, A. and Joshi, O. (2020). The continuing practice of untouchability in india: Patterns and mitigating influences. *Economic & Political Weekly*, 55(2):37–45.

A Measurement, Samples, and Spatial Construction

This section defines the empirical objects used in the paper and explains how they are constructed. The design keeps three objects distinct: the stock of active public primary schools, residents' feasible access to those schools, and the allocation of enrolled pupils across the local public–private school market. The first two objects define the policy-induced change in access. The third defines the caste-segregation outcome.

A.1 Administrative school records and the active school stock

The school panel combines annual DISE and UDISE records for primary schooling. Schools are harmonised across years using stable school identifiers, management categories, grade spans, enrolment variables, and geocoded locations. Publicly managed primary schools form the supply used to construct feasible public-school access. The segregation market is broader: it includes recognised public and private primary schools with the enrolment information required to measure caste composition in a given year.

For the access measures, I use an active public-school stock. A school's entry year is based on the reported primary opening year where available and the reported general opening year otherwise. Once a public primary school enters the stock, temporary gaps in annual reporting do not make it disappear from residents' feasible menu. This convention treats access as physical availability rather than as a measure of annual reporting. The segregation outcome necessarily uses annual records because contemporaneous caste-specific enrolment is required to measure how pupils are distributed across schools.

A.2 Resident-centred public-school access

The unit of analysis is the Census village, but village boundaries are not treated as school catchments. I locate residents within each village using inhabited cells from the 100-metre GHSL population grid. For each inhabited cell c and year t, I count the active public primary schools within one kilometre of the cell centroid:

$$N_{ct} = \sum_{s \in \mathcal{A}_t} \mathbf{1}\{\text{dist}(c, s) \leq 1{,}000\}, \tag{14}$$

where $\mathcal{A}_t$ is the active public-primary-school stock in year t. Schools are counted regardless of whether they lie inside the focal village. A school across an administrative boundary therefore enters a resident's menu whenever it falls within the one-kilometre radius.

Let $\mathcal{C}_v$ denote the inhabited cells in village v and p_c the fixed population weight of cell c. The share of residents with at least k nearby public schools is

$$A_{vt}^{(k)} = \frac{\sum_{c \in \mathcal{C}_v} p_c \mathbf{1}\{N_{ct} \geq k\}}{\sum_{c \in \mathcal{C}_v} p_c}. \tag{15}$$

The main multiple-school access measure is $A_{vt}^{(2)}$, the share of residents with at least two public primary schools within one kilometre. The population weights are held fixed at their 2010

distribution, so changes in measured access reflect changes in nearby public-school supply rather than changes in the within-village location of residents.

The cumulative access measures can be complemented by mutually exclusive menu states. For $j \geq 0$, define

$$A_{vt}^{(=j)} = \frac{\sum_{c \in \mathcal{C}_v} p_c \mathbf{1}\{N_{ct} = j\}}{\sum_{c \in \mathcal{C}_v} p_c}. \tag{16}$$

The analysis distinguishes residents with no nearby public school, exactly one, exactly two, exactly three, exactly four, and five or more. Movement from exactly one school to two or more creates a choice margin that is conceptually different from movement from no school to first access.

A.3 Construction of baseline eligibility

I construct baseline eligibility by applying the one-kilometre access rule to the 2008 active public-primary-school stock and the inhabited population cells described above. A village is eligible if at least one inhabited cell lies more than one kilometre from its nearest active public primary school. Distances are measured as straight-line distances in the same projected coordinate system used for the access measures, and eligibility is held fixed throughout the panel.

The same 2008 spatial construction also identifies baseline target cells: inhabited cells lying more than one kilometre from the nearest active public primary school. I use these target cells in the cross-village exposure check in Appendix C.2 to identify ineligible villages whose residents' one-kilometre school markets overlap areas targeted by the access rule. This overlap measure is a separate exposure diagnostic; it does not redefine baseline eligibility.

A.4 The local public–private school market

Pupils' residential locations are not observed in the administrative school records. The segregation market is therefore defined around the village polygon rather than around individual resident cells. Let P_v denote the polygon for village v. The annual local market contains recognised primary schools with the required enrolment information that lie inside or within one kilometre of the polygon:

$$\mathcal{S}_{vt} = \{s : \text{dist}(s, P_v) \leq 1{,}000\}. \tag{17}$$

The market includes public and private schools. This is deliberate: an expansion of public provision can change enrolment both within the public sector and across sectors. A public-school-only outcome would miss part of that local allocation response.

The different spatial anchors reflect the information available in the two datasets. Resident-centred access measures which public schools are geographically feasible from where people live. The polygon-centred market measures how enrolled SC and non-SC pupils are distributed across the schools serving the local area.

A.5 Caste dissimilarity and the random-allocation benchmark

Let SC_{st} and NSC_{st} denote Scheduled Caste and non-Scheduled-Caste primary enrolment in school s in year t. For active schools $s \in \mathcal{S}_{vt}$, observed dissimilarity is

$$D_{vt}^{\text{obs}} = \frac{1}{2} \sum_{s \in \mathcal{S}_{vt}} \left| \frac{SC_{st}}{SC_{vt}} - \frac{NSC_{st}}{NSC_{vt}} \right|, \tag{18}$$

where $SC_{vt} = \sum_s SC_{st}$ and $NSC_{vt} = \sum_s NSC_{st}$. The index is zero when the two groups are distributed proportionally across local schools and rises as their school distributions diverge.

Some unevenness arises mechanically when finite numbers of pupils are divided across several schools. I therefore construct an expected-random benchmark. For each village–year, the simulation holds fixed the realised school set, each school's caste-valid enrolment, and the market totals of SC and non-SC pupils. It randomly reallocates caste labels across the fixed school places without replacement and recalculates dissimilarity. The benchmark is the mean across 100 allocations:

$$D_{vt}^{\text{rand}} = \frac{1}{100} \sum_{r=1}^{100} D_{vt}^{(r)}. \tag{19}$$

The beyond-random component is

$$D_{vt}^{\text{beyond}} = D_{vt}^{\text{obs}} - D_{vt}^{\text{rand}}. \tag{20}$$

The headline table estimates all three outcomes on identical observations. This decomposition distinguishes a change in systematic caste sorting from a change in mechanical unevenness generated by market size and school sizes.

A.6 The two empirical margins and the predetermined market sample

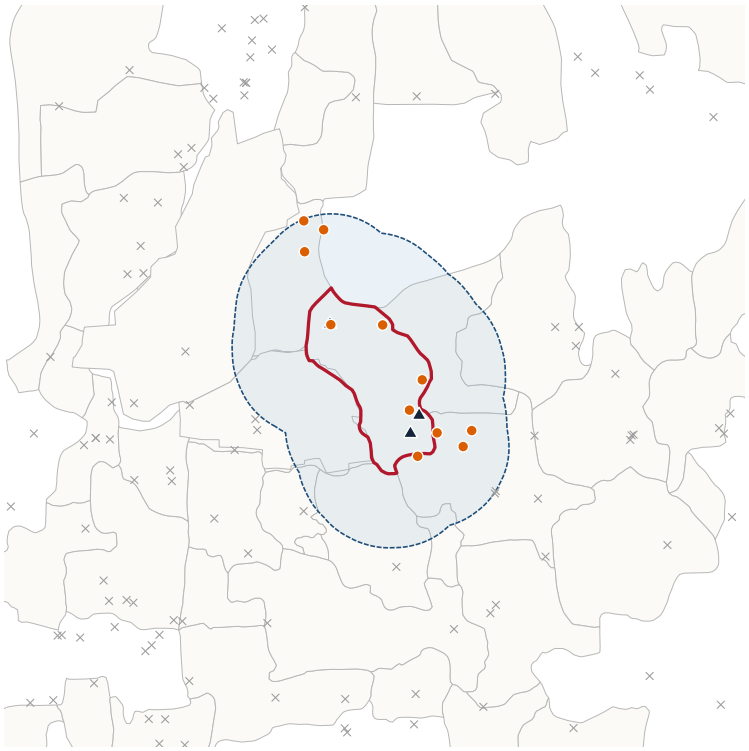
Margin 1 uses the broad policy sample to estimate how baseline eligibility changes feasible public-school access. Its outcomes are population-weighted resident access measures, not school counts at the village centroid and not individual pupil transitions.

Margin 2 asks how caste allocation changes when feasible public-school choice expands in places where an allocation across schools was already observable before the reform. The predetermined established-market sample requires at least two active public or private primary schools within one kilometre in both 2006 and 2007. The restriction does not use the 2008 school count, post-reform school entry, or post-reform segregation. It therefore separates the sorting response from the mechanical creation of a first multi-school market. In the IV specifications, $A_{vt}^{(2)}$ is scaled so that coefficients represent a 10-percentage-point increase in multiple-school access.

A.7 Spatial reach across administrative boundaries

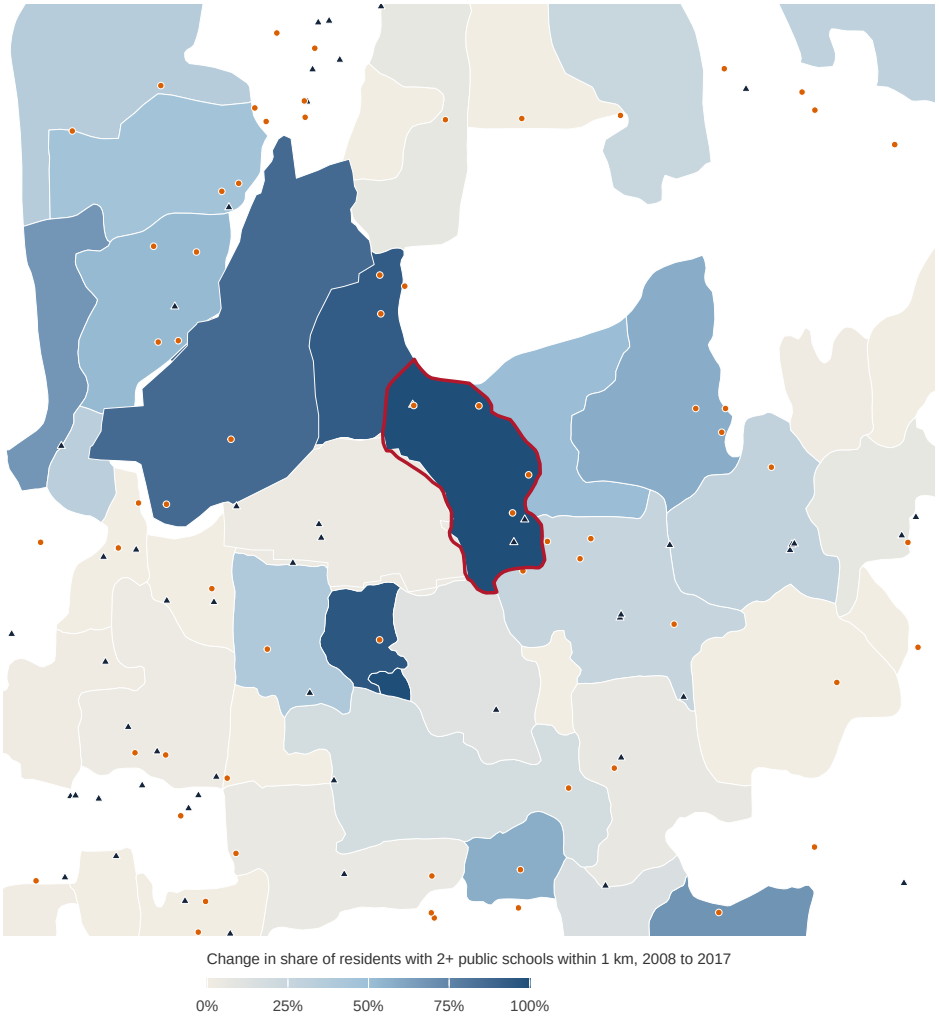
Figure A1 illustrates why the access measure is built outward from inhabited cells. Figure A2 then places the same village in its surrounding ten-kilometre window and shows that changes in access need not stop at administrative borders.

Figure A1: Public schools within one kilometre of a village's residents



Notes: The figure shows the same village as Figure 1, outlined in red, with neighbouring Census villages shown in outline. The shaded region bounded by the dashed line is the area within one kilometre of the village boundary. Public schools inside that region are shown as filled symbols: navy triangles denote schools open before 2010 and orange circles denote schools opened between 2010 and 2017. Grey crosses denote public schools beyond the region. Filled symbols outside the village boundary illustrate that residents near the boundary may have nearby schools in a neighbouring village. The resident-centred access measure counts such schools because it is computed outward from inhabited population cells. The figure is illustrative and is not a sample statistic.

Figure A2: Changes in public-school access across neighbouring Census villages



Notes: The figure shows Census villages in a ten-kilometre window centred on the village illustrated in Figure 1, which is outlined in red. Shading gives the change in $A_{vt}^{(2)}$ between 2008 and 2017. Only villages in the estimating sample are drawn; blank areas within the window are villages outside the sample. Navy triangles denote public schools open before 2010, and orange circles denote schools opened between 2010 and 2017. Both cohorts are shown irrespective of the village in which they are located. The figure is illustrative and is not a sample statistic.

A.8 Why school counts, access, and segregation are kept separate

The active school stock counts physical public-school supply. Resident-centred access measures the population share for whom that supply is within a feasible one-kilometre radius. Segregation measures the distribution of enrolled SC and non-SC pupils across the full local public–private school market. A new school can increase the stock without changing access for many residents; access can change without a corresponding change in aggregate enrolment; and an unchanged market-level caste share can coexist with substantial reallocation across schools.

The supplementary exercises address distinct alternatives to the paper's interpretation. The random-allocation benchmark isolates systematic sorting from finite-sample unevenness. The nearest-school benchmark assesses how much of the response can be reproduced by residential geography and distance. Market-composition and reporting checks examine whether the measured effect is driven by changes in the population observed in the local market. Incumbent-school and input analyses examine whether the response is accompanied by differentiation within the pre-existing public-school network.

B Robustness to Timing, Sample, and Outcome Definitions

This section varies the treatment of the transition year, the definition of a predetermined established market, and the segregation index. It then verifies that the results are not explained by changes in caste-enrolment reporting or by the disappearance of either caste group from the measured market.

B.1 Treatment of the 2009 transition year

The preferred pooled specification excludes 2009 because the Right to Education Act was enacted in that year but entered formal implementation in 2010. Table A1 alternatively treats 2009 as part of the pre-implementation period or as the first post-reform year. The first stage and the segregation estimates remain close to the preferred specification.

B.2 Alternative definitions of an established school market

The headline sample requires at least two active public or private primary schools in both 2006 and 2007. Table A2 strengthens that predetermined restriction using the 2008 school count. The estimates are stable across these alternative definitions.

B.3 Alternative segregation indices

Dissimilarity is the main outcome because it measures the evenness of the SC and non-SC distributions across schools. Table A3 shows that the result is not specific to this index by reporting Theil's entropy index and SC isolation on the same established-market sample.

B.4 Reporting and outcome support

The reform could affect which schools report caste enrolment or whether both caste groups are represented in the local market. Table A4 examines these margins directly in the predetermined established-market sample. Reporting is nearly complete, and the estimates provide no evidence that changes in reporting account for the segregation result. Any estimated change in the presence of both caste groups is very small in magnitude.

Table A1: Alternative treatment of the 2009 transition year

	2009 excluded (1)	2009 as pre (2)	2009 as post (3)
A. First stage			
Eligibility × Post	0.3253***	0.3230***	0.2822***
	(0.0543)	(0.0510)	(0.0491)
First-stage F statistic	35.9	40.1	33.1
B. Reduced form			
Observed dissimilarity	0.0054***	0.0050***	0.0049***
	(0.0008)	(0.0008)	(0.0008)
Expected under random allocation	-0.0002	-0.0002	-0.0002
	(0.0003)	(0.0003)	(0.0003)
Beyond-random dissimilarity	0.0057***	0.0052***	0.0051***
	(0.0008)	(0.0007)	(0.0007)
C. Two-stage least squares			
Observed dissimilarity	0.0167***	0.0155***	0.0174***
	(0.0023)	(0.0021)	(0.0025)
Expected under random allocation	-0.0007	-0.0007	-0.0007
	(0.0009)	(0.0009)	(0.0010)
Beyond-random dissimilarity	0.0174***	0.0162***	0.0181***
	(0.0025)	(0.0023)	(0.0027)
Villages	350,331	350,738	350,738
Observations	3,185,709	3,507,294	3,507,294

Notes: Column 1 excludes 2009, as in the preferred specification. Column 2 retains 2009 as part of the pre-implementation period. Column 3 treats 2009 as the first post-reform year. The sample, outcomes, controls, fixed effects, scaling, and inference are otherwise identical to the headline established-market specification. Standard errors are clustered by district.

Table A2: Robustness to alternative established-market samples

	Main 2006–07 (1)	Strict-plus (2)	Strict 2006–08 (3)
A. First stage			
Eligibility × Post	0.3253***	0.3252***	0.3244***
	(0.0543)	(0.0542)	(0.0544)
First-stage F statistic	35.9	36.0	35.6
B. Reduced form			
Observed dissimilarity	0.0054***	0.0053***	0.0052***
	(0.0008)	(0.0008)	(0.0008)
Expected under random allocation	-0.0002	-0.0003	-0.0003
	(0.0003)	(0.0003)	(0.0003)
Beyond-random dissimilarity	0.0057***	0.0055***	0.0055***
	(0.0008)	(0.0007)	(0.0007)
C. Two-stage least squares			
Observed dissimilarity	0.0167***	0.0162***	0.0161***
	(0.0023)	(0.0023)	(0.0023)
Expected under random allocation	-0.0007	-0.0008	-0.0009
	(0.0009)	(0.0009)	(0.0009)
Beyond-random dissimilarity	0.0174***	0.0170***	0.0170***
	(0.0025)	(0.0025)	(0.0025)
Villages	350,331	349,484	348,288
Observations	3,185,709	3,177,941	3,168,749

Notes: Column 1 reports the headline established-market sample, which requires at least two active public or private primary schools within one kilometre in both 2006 and 2007. Column 2 additionally requires an observed 2008 count, where available, to indicate at least two schools; villages with a missing 2008 count remain in the sample. Column 3 requires observed counts of at least two schools in 2006, 2007, and 2008. All columns exclude 2009 and otherwise follow the headline specification. Standard errors are clustered by district.

Table A3: Alternative segregation indices

	Observed D (1)	Theil H (2)	SC isolation (3)
A. First stage			
Eligibility × Post	0.3253***	0.3253***	0.3253***
	(0.0543)	(0.0543)	(0.0543)
First-stage F statistic	35.9	35.9	35.9
B. Reduced form			
Eligibility × Post	0.0054***	0.0040***	0.0019***
	(0.0008)	(0.0007)	(0.0007)
C. Two-stage least squares			
Access to 2+ public schools (10 p.p.)	0.0167***	0.0122***	0.0060***
	(0.0023)	(0.0018)	(0.0019)
	[0.0122, 0.0212]	[0.0087, 0.0157]	[0.0022, 0.0097]
Mean of outcome	0.431	0.205	0.374
Villages	350,331	350,331	350,331
District clusters	554	554	554
Observations	3,185,709	3,185,709	3,185,709

Notes: All columns use a common established-market sample requiring non-missing observed dissimilarity, Theil's entropy index, and the Scheduled Caste isolation index. The endogenous variable is the share of residents with at least two public primary schools within one kilometre, scaled so that the 2SLS coefficient represents a 10-percentage-point increase. The remaining specification and inference follow the headline table. The transition year 2009 is excluded. Standard errors are clustered by district.

Table A4: Effects on reporting and the presence of both caste groups

	Reporting schools (1)	Reporting share (2)	Both groups present (3)
Reduced form	0.0162	-0.0009	-0.0211***
	(0.0284)	(0.0008)	(0.0061)
2SLS: 2+ access (10 p.p.)	0.0496	-0.0029	-0.0649***
	(0.0876)	(0.0025)	(0.0224)
Mean of outcome	6.462	99.989	99.953
Villages		350,331	
Observations		3,185,709	

Notes: Column 1 reports the number of schools contributing valid caste-enrolment information to the local market. Column 2 reports the share of active schools with valid reporting, in percentage points. Column 3 reports the percentage of village–year markets in which both SC and non-SC pupils are represented. Reduced-form rows report coefficients on baseline eligibility interacted with the post-reform period. The 2SLS rows instrument multiple- school access with the same interaction; coefficients represent a 10-percentage-point increase in access. The transition year 2009 is excluded. All specifications include village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls. Standard errors are clustered by district.

C Identification and Dynamic Checks

This section examines the identifying comparisons behind the pooled estimates. It first removes the Census-2011 characteristic-by-year controls and then restricts the control group using pre-reform spatial overlap. It next presents the timing of the first stage and reduced form using 2008 as the event-study reference year. The final checks allow eligible villages to follow a different linear path and apply relative-magnitude restrictions to possible violations of parallel trends. The reduced form remains the policy estimand; 2SLS scales that effect by the eligibility-induced change in feasible choice.

C.1 Audit of the Census-2011 characteristic-by-year controls

The preferred specifications interact year indicators with time-invariant village characteristics recorded in the 2011 Census: log population, log area, SC and ST population shares, all-weather-road access, bus access, mobile coverage, digital connectivity, a public distribution shop, and an ATM. Because implementation began before these characteristics were recorded, Table A5 removes all of the interactions while holding the estimation sample fixed. The first-stage, reduced-form, and 2SLS estimates are nearly unchanged. The controls improve precision and allow paths to differ across observed village types, but the identifying conclusion does not depend on them.

C.2 Cross-village exposure and clean controls

A village's feasible school market can cross an administrative boundary. Some baseline-ineligible villages may therefore be close enough to schools created for nearby targeted residents to be indirectly exposed to the expansion. The clean-control exercise retains every baseline-eligible village but restricts controls to baseline-ineligible villages whose residents' one-kilometre school markets have zero overlap with any 2008 target cell. The overlap measure is constructed directly from resident cells rather than from village-centroid proximity.

Table A6 shows that excluding potentially exposed controls does not attenuate the segregation result. This exercise uses pre-reform geography to sharpen the comparison group; it does not redefine treatment using a post-reform outcome.

C.3 Access dynamics and the implementation break

Figure A3 reports the unrestricted access dynamics with 2008 as the omitted year and 2009 retained as a transition-year estimate. Eligible villages were on a different pre-policy path in access to multiple public schools. The direction of this path reverses after enactment and implementation. Figure A4 makes that break explicit, and Figure A5 reports the implied effects under exact continuation of the estimated pre-policy differential trend.

C.4 Sorting dynamics with 2008 as the reference year

The pooled specification excludes 2009 because it lies between enactment and implementation. Using 2009 as the omitted event-study year would, however, absorb any anticipation or

Table A5: Audit of the Census-2011 characteristic-by-year interactions

	Full Census-2011 controls (1)	No Census-2011 controls (2)
A. First stage		
Eligibility × Post	0.3253***	0.3230***
	(0.0543)	(0.0516)
First-stage F statistic	35.95	39.25
B. Reduced form		
Observed dissimilarity	0.0054***	0.0055***
	(0.0008)	(0.0008)
Expected under random allocation	-0.0002	-0.0000
	(0.0003)	(0.0003)
Beyond-random dissimilarity	0.0057***	0.0056***
	(0.0008)	(0.0007)
C. Two-stage least squares		
Observed dissimilarity	0.0167***	0.0172***
	(0.0023)	(0.0025)
Expected under random allocation	-0.0007	-0.0000
	(0.0009)	(0.0010)
Beyond-random dissimilarity	0.0174***	0.0172***
	(0.0025)	(0.0026)
Villages	350,331	350,331
District clusters	554	554
Observations	3,185,709	3,185,709

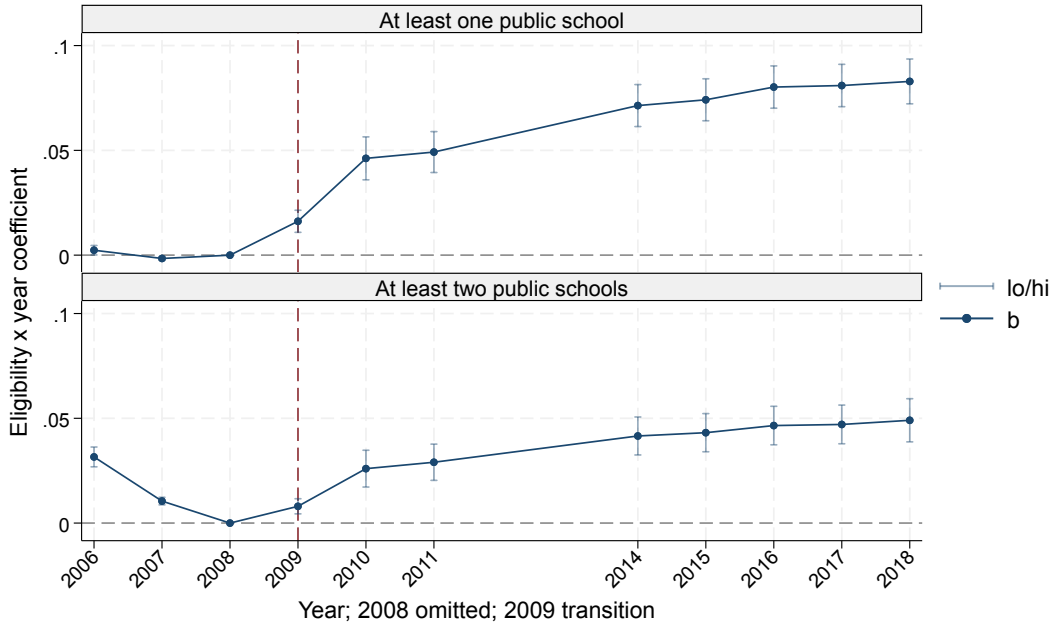
Notes: The unit of observation is a village–year in the predetermined established-market sample; 2009 is excluded. Column 1 reports the preferred specification with Census-2011 characteristics interacted with year. Column 2 removes all Census-2011 interactions while holding the sample fixed. Panel A reports the first stage, Panel B the policy reduced form, and Panel C the 2SLS scaling. All specifications include village fixed effects and district-by-year fixed effects. The access variable is expressed in units of 10 percentage points. Standard errors are clustered by district and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6: Robustness to excluding overlapping control markets

	All controls (1)	Excluding exposed controls (2)
A. First stage		
Eligibility × Post	0.3253***	0.5396***
	(0.0543)	(0.0822)
First-stage F statistic	35.9	43.1
B. Reduced form		
Observed dissimilarity	0.0054***	0.0085***
	(0.0008)	(0.0014)
Expected under random allocation	-0.0002	-0.0001
	(0.0003)	(0.0005)
Beyond-random dissimilarity	0.0057***	0.0086***
	(0.0008)	(0.0012)
C. Two-stage least squares		
Observed dissimilarity	0.0167***	0.0158***
	(0.0023)	(0.0020)
Expected under random allocation	-0.0007	-0.0002
	(0.0009)	(0.0009)
Beyond-random dissimilarity	0.0174***	0.0159***
	(0.0025)	(0.0021)
Share of control villages retained	1.000	0.418
Villages	350,331	261,177
Observations	3,185,709	2,376,688

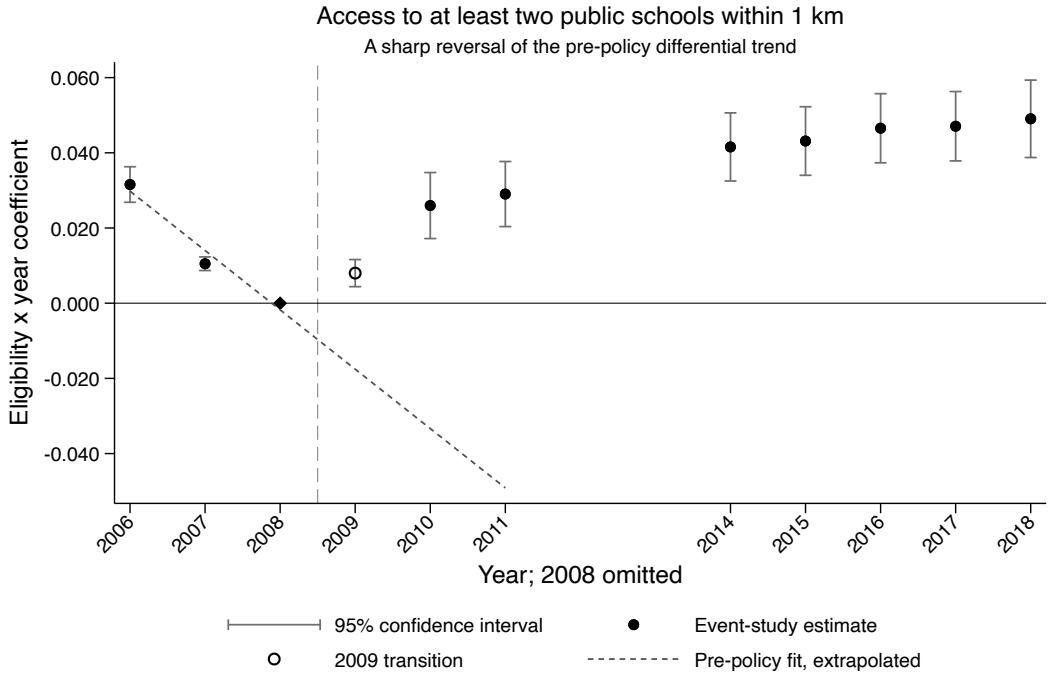
Notes: Column 1 uses the established-market headline sample. Column 2 uses the clean-control specification that retains all baseline-eligible villages and restricts controls to baseline-ineligible villages whose residents' one-kilometre school markets have zero overlap with any 2008 target cell. The overlap measure is constructed from the final resident-cell target-overlap file. All three segregation outcomes use identical observations within each column. The transition year 2009 is excluded; all other specification and inference choices follow the headline table. Standard errors are clustered by district.

Figure A3: Eligibility and nearby public-school access, with 2008 as the reference year



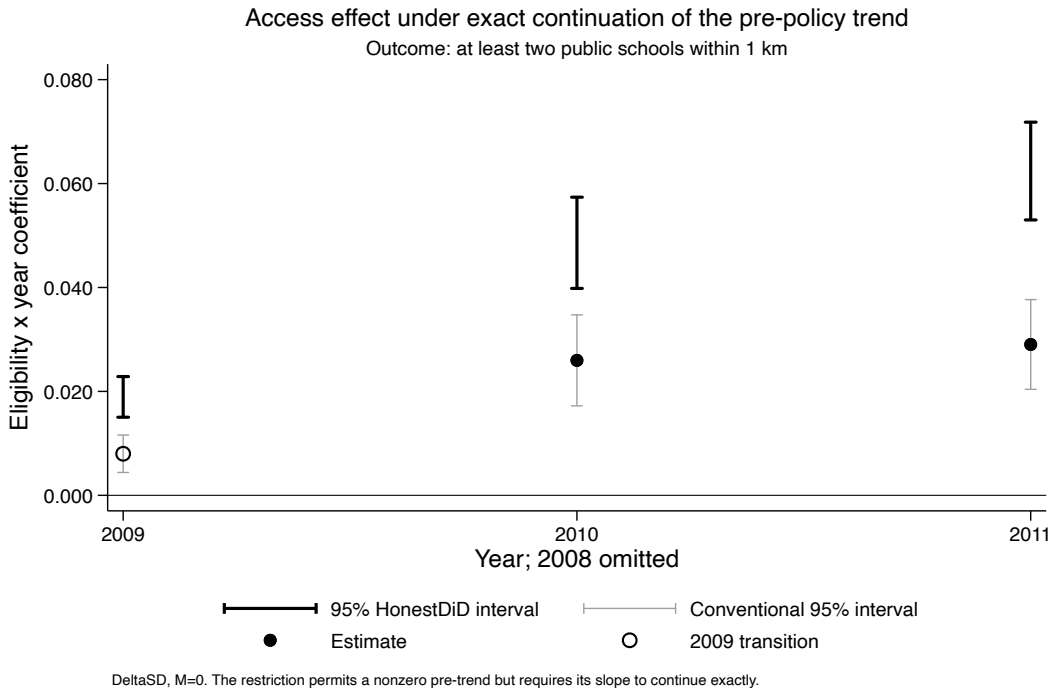
Notes: The figure reports coefficients on baseline eligibility interacted with year indicators in the broad policy sample. The omitted year is 2008, and 2009 is marked as a transition year. The upper panel reports the share of residents with at least one public primary school within one kilometre; the lower panel reports the share with at least two. Bars are 95 percent confidence intervals based on standard errors clustered by district. All specifications include village fixed effects, district-by-year fixed effects, and Census-2011 village characteristics interacted with year.

Figure A4: Trend break in access to at least two nearby public schools



Notes: Points are eligibility-by-year coefficients for the share of residents with at least two public primary schools within one kilometre, with 2008 omitted. Bars are 95 percent confidence intervals based on standard errors clustered by district. The dashed line extends the estimated differential pre-policy trend. The reversal after enactment and implementation separates the policy-induced access change from continuation of the earlier path.

Figure A5: Access effects under exact continuation of the pre-policy trend



Notes: Points are eligibility-by-year coefficients for access to at least two public primary schools, with 2008 omitted. The narrower intervals are conventional 95 percent confidence intervals. The wider intervals incorporate uncertainty while imposing exact continuation of the estimated pre-policy differential trend. The 2009 coefficient is a transition-period estimate. Standard errors are clustered by district.

transition-period response into the reference category. Figure A6 therefore omits 2008, the year in which eligibility is fixed, and estimates a separate coefficient for 2009. The pre-period coefficients provide no evidence of an upward differential trend in observed or beyond-random dissimilarity.

C.5 Exact continuation of the sorting pretrend

Figure A7 anchors the counterfactual to exact continuation of the estimated pre-period path. Because the measured differential pretrends in observed and beyond-random dissimilarity are negative, accounting for their continuation strengthens rather than removes the positive implementation break. Expected dissimilarity under random allocation remains close to zero.

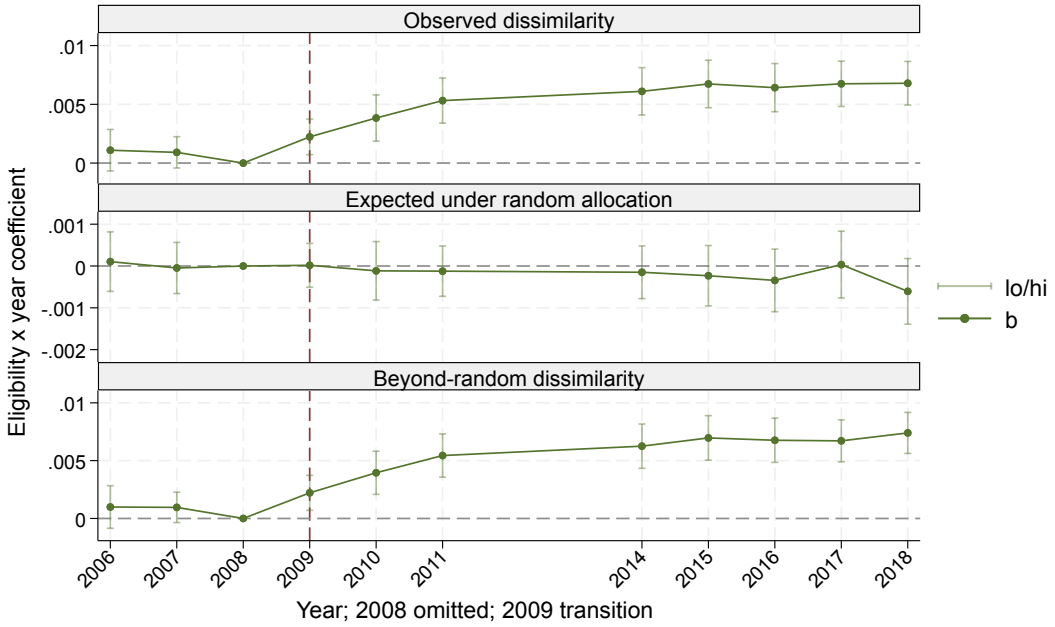
C.6 Eligible-specific trends and reduced-form/IV scaling

Table A7 adds an eligible-specific linear calendar- time trend to the pooled first-stage, reduced-form, and 2SLS specifications. This allows eligible villages to remain on a different linear path throughout the estimation window. The reduced-form effects remain positive for observed and beyond-random dissimilarity, while the random-allocation component remains close to zero. The first stage becomes smaller but remains statistically distinguishable from zero.

C.7 Relative-magnitude sensitivity analysis

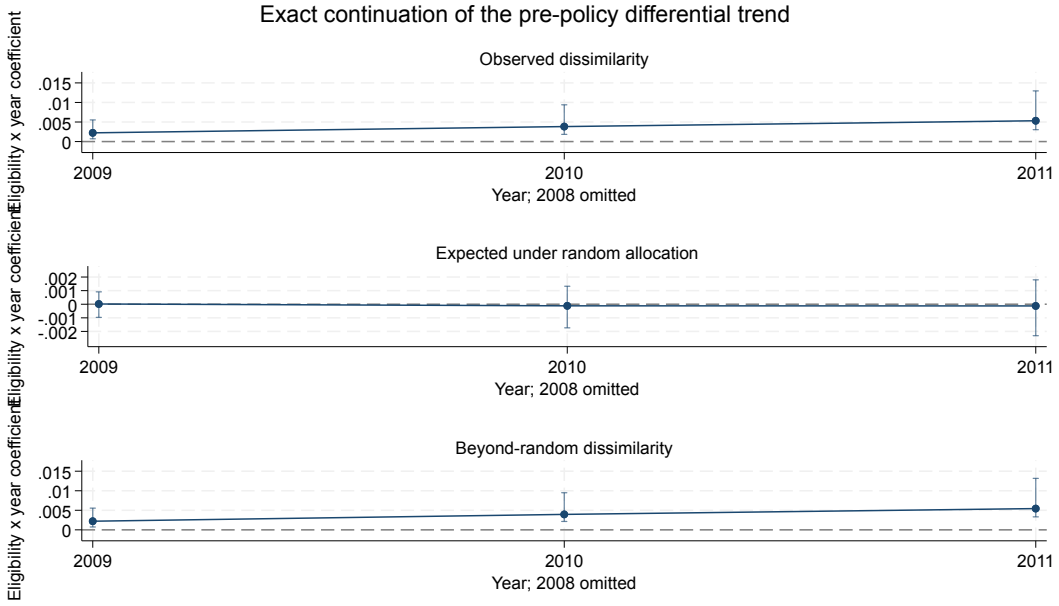
Table A8 applies the relative-magnitude restrictions of Rambachan and Roth (2023) to the reduced-form event-study coefficients. The early estimand averages 2010 and 2011, while the longer-run estimand averages 2014–2018. Figures A8 and A9 show how the robust confidence sets change over the evaluated grid of $\bar{M}$ values.

Figure A6: Eligibility and caste sorting, with 2008 as the reference year



Notes: The figure reports coefficients on baseline eligibility interacted with year indicators. The omitted year is 2008. The 2009 coefficient is retained and marked as a transition-period estimate because the Act was enacted in August 2009 and implemented in April 2010. The three panels report observed dissimilarity, expected dissimilarity under random allocation, and their difference. Bars are 95 percent confidence intervals based on standard errors clustered by district. All specifications include village fixed effects, district-by-year fixed effects, and Census-2011 village characteristics interacted with year.

Figure A7: Early sorting effects under exact continuation of the pre-policy trend



Notes: The three panels report observed dissimilarity, expected dissimilarity under random allocation, and beyond-random dissimilarity. Points are eligibility-by-year coefficients from the event study with 2008 omitted. The narrower intervals are conventional 95 percent confidence intervals. The wider intervals incorporate uncertainty while imposing exact continuation of the estimated pre-policy differential trend. The 2009 coefficient is a transition-period estimate. Standard errors are clustered by district.

Table A7: Policy effects, induced feasible choice, and eligible-specific trends

	Baseline (1)	Eligible-specific trend (2)
A. Policy reduced form: central estimates		
Observed dissimilarity	0.0054***	0.0033***
	(0.0008)	(0.0011)
Expected under random allocation	-0.0002	0.0000
	(0.0003)	(0.0004)
Beyond-random dissimilarity	0.0057***	0.0033***
	(0.0008)	(0.0010)
B. First stage: induced feasible choice		
Eligibility × Post	0.3253***	0.1716***
	(0.0543)	(0.0505)
First-stage F statistic	35.95	11.54
C. 2SLS scaling through induced 2+ access		
Observed dissimilarity	0.0167***	0.0192***
	(0.0023)	(0.0057)
Expected under random allocation	-0.0007	0.0000
	(0.0009)	(0.0022)
Beyond-random dissimilarity	0.0174***	0.0192***
	(0.0025)	(0.0057)
Villages	350,331	350,331
District clusters	554	554
Observations	3,185,709	3,185,709

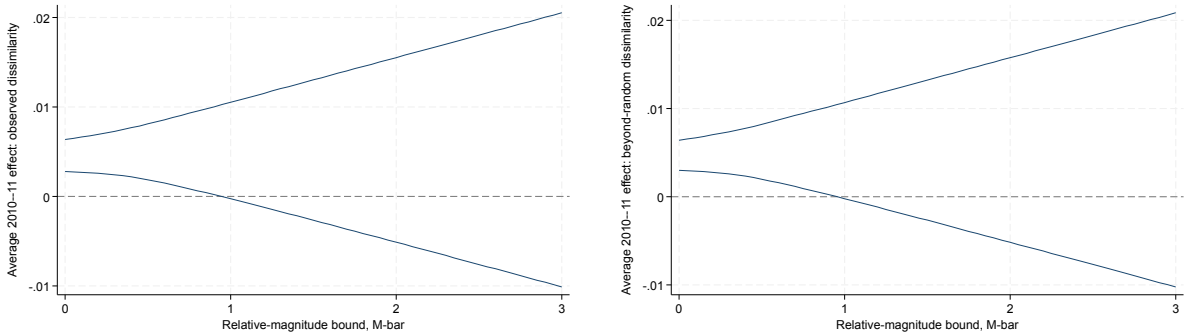
Notes: The unit of observation is a village–year in the predetermined established-market sample; 2009 is excluded. Column 1 reports the preferred specification. Column 2 adds baseline eligibility interacted with a linear calendar-time trend to each equation. Panel A reports the policy reduced form, Panel B the first stage for access to at least two nearby public primary schools, and Panel C the corresponding 2SLS scaling. Both columns include village fixed effects, district-by-year fixed effects, and Census-2011 village characteristics interacted with year. The access variable is expressed in units of 10 percentage points. Standard errors are clustered by district and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A8: Relative-magnitude sensitivity of the reduced-form sorting effects

	Observed dissimilarity (1)	Beyond-random dissimilarity (2)
A. Early implementation: average 2010–11		
Conventional estimate	0.0046	0.0047
Standard error	(0.0009)	(0.0009)
Conventional 95% CI	[0.0028, 0.0064]	[0.0030, 0.0064]
Robust CI, $\bar{M} = 0.5$	[0.0019, 0.0081]	[0.0020, 0.0082]
Robust CI, $\bar{M} = 1$	[-0.0003, 0.0105]	[-0.0002, 0.0107]
First $\bar{M}$ including zero	0.95	1.00
B. Longer run: average 2014–18		
Conventional estimate	0.0066	0.0068
Standard error	(0.0009)	(0.0009)
Conventional 95% CI	[0.0048, 0.0084]	[0.0051, 0.0085]
Robust CI, $\bar{M} = 0.5$	[0.0006, 0.0135]	[0.0009, 0.0138]
Robust CI, $\bar{M} = 1$	[-0.0053, 0.0194]	[-0.0052, 0.0199]
First $\bar{M}$ including zero	0.60	0.60

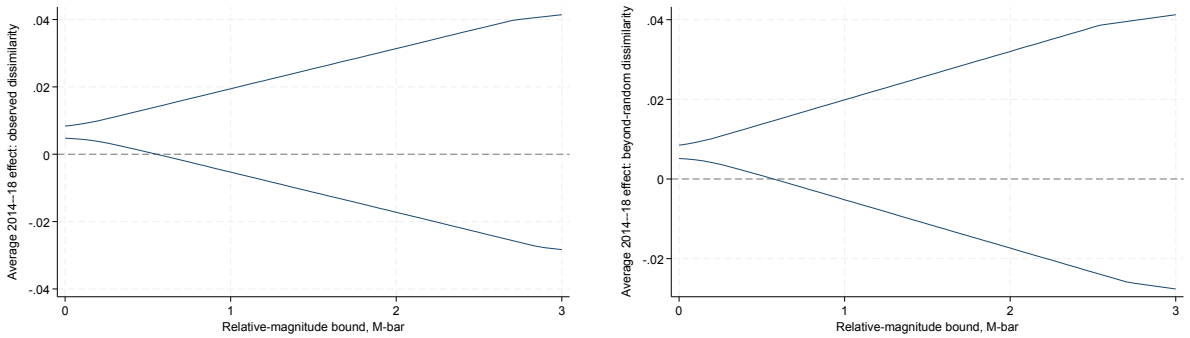
Notes: The analysis is applied to reduced-form event-study coefficients with 2008 omitted. The early estimand averages 2010 and 2011; the longer-run estimand averages 2014–2018. Robust confidence sets bound the magnitude of post-treatment violations of parallel trends relative to the largest pre-treatment change, as indexed by $\bar{M}$. "First $\bar{M}$ including zero" reports the first value on the evaluated grid at which the robust confidence set contains zero. Conventional intervals use standard errors clustered by district.

Figure A8: Relative-magnitude sensitivity: average 2010–11 sorting effect



Notes: The left panel reports the average 2010–11 reduced-form effect on observed dissimilarity; the right panel reports beyond-random dissimilarity. The solid line is the point estimate, and the shaded region is the robust confidence set at each value of $\bar{M}$. The horizontal line marks zero.

Figure A9: Relative-magnitude sensitivity: average 2014–18 sorting effect



Notes: The left panel reports the average 2014–18 reduced-form effect on observed dissimilarity; the right panel reports beyond-random dissimilarity. The solid line is the point estimate, and the shaded region is the robust confidence set at each value of $\bar{M}$. The horizontal line marks zero.

D Cross-Caste Exposure: Measurement and Sensitivity

Section 5.3 relates changes in school allocation to the schoolmates experienced by SC and non-SC pupils. This appendix sets out the exposure measures, the exact random-allocation benchmark, and the evidence used to choose the main sample. It also reports how the results change with the definition of low contact. The purpose is to distinguish changes in school-level exposure from mechanical features of the outcomes and from assumptions about the counterfactual path.

D.1 Pupil exposure and the random-allocation benchmark

Let n_{svt} be primary enrolment in school s in village v's local market in year t, and let n_{gsvt} be enrolment of group g, where g is SC or non-SC. Write $N_{gvt} = \sum_s n_{gsvt}$ and $N_{vt} = \sum_s n_{svt}$ for the corresponding market totals. The out-group schoolmate count, out-group enrolment share, and zero-contact rate are

$$C_{gvt} = \sum_s \frac{n_{gsvt}}{N_{gvt}} n_{-g,svt}, \tag{21}$$

$$S_{gvt} = 100 \sum_s \frac{n_{gsvt}}{N_{gvt}} \frac{n_{-g,svt}}{n_{svt}}, \tag{22}$$

$$Z_{gvt} = 100 \sum_s \frac{n_{gsvt}}{N_{gvt}} \mathbf{1}\{n_{-g,svt} = 0\}. \tag{23}$$

These measures weight schools by the number of pupils from the focal group. The share denominator is total primary enrolment, including the focal pupil. Own-group schoolmate counts exclude that pupil, so own-group plus out-group schoolmates equal school size minus one. With the same observations and specification, the estimated effects on the two counts therefore sum to the effect on school size. This is an accounting identity, not an additional mechanism test. The regression unit remains the village–year, and the reported means average these within-market measures equally across observed villages in 2008.

For low-contact outcomes, observed separation is compared with random allocation of caste labels across the realised school places. Conditional on a focal pupil being from group g, the number of out-group pupils among the other $n_{svt} - 1$ places in school s has a hypergeometric distribution:

$$K_{gsvt} \sim \text{Hypergeom}(N_{vt} - 1,\ N_{-g,vt},\ n_{svt} - 1). \tag{24}$$

The arguments are the remaining population, the number of out-group pupils in that population, and the number of remaining school places. For an out-group count cutoff k, the expected pupil-weighted low-contact rate is

$$B_{gvt}(k) = 100 \sum_s \frac{n_{svt}}{N_{vt}} \Pr(K_{gsvt} \le k). \tag{25}$$

Under random allocation, a group-g pupil attends school s with probability equal to that school's share of market enrolment, n_{svt}/N_{vt}. In the observed outcome, the weight is instead that school's share of group-g enrolment, n_{gsvt}/N_{gvt}. The different weights are intentional: random allocation changes where pupils from each group attend school as well as the composition of the school they attend. Exact zero uses $k = 0$. For an out-group enrolment-share threshold q, the school-

specific integer cutoff is $\lfloor qn_{svt} \rfloor$, applied consistently to observed and random outcomes. Excess low contact is observed low contact minus B.

Conditioning on the focal pupil matters most in small schools. A random draw over all school places would not condition on that pupil's known group and need not reproduce the outcome in a one-school market. The conditional benchmark does: with one school, the realised group totals exhaust the market totals, so observed and expected low contact coincide. There is no within-market allocation margin for the benchmark to distinguish in that case.

In the established-market non-SC sample, the mean observed zero-contact rate is 19.2444 percent and the excess mean is 16.5772 percentage points. Their difference, 2.6672 percent, is the mean random-allocation rate. For SC pupils the corresponding random rate is 0.0173 percent, compared with an observed rate of 2.85 percent. Essentially all of the observed SC zero-contact rate is therefore in excess of what random allocation would produce; this does not identify whether geography, preferences or other constraints explain that separation. These are calculated using the same samples and weights as the observed outcomes. They describe the baseline distribution, not effects of eligibility. Nor is random assignment across all places a feasible school-siting or pupil-assignment policy: the benchmark does not preserve each child's travel distance or household choices. It isolates the arithmetic associated with market caste totals and school sizes.

D.2 The established-market and broader contact samples

The main contact table applies the same baseline established-market restriction as the segregation analysis: at least two active recognised primary schools in both 2006 and 2007. It additionally requires both groups in the market in 2008. Within each group, observations must have positive focal-group enrolment and complete outcomes and regressors. The resulting contact samples contain 335,709 villages for SC pupils and 335,824 for non-SC pupils. They share a baseline market definition with the headline segregation sample but not its exact set of regression observations.

The broader sample instead requires at least one active school in both 2006 and 2007, retaining the same 2008 group-presence condition. It includes markets in which the reform could create a multi-school allocation margin. This is a legitimate population for contact outcomes, which remain defined in single-school markets, but it is not simply a larger version of the established-market estimand. Neither sample selects villages on their post-reform number of schools, although current outcome availability and focal-group enrolment can affect the observed panel.

Table A9 reports the broader-sample estimates. Per ten-percentage-point increase in access, out-group schoolmate counts fall by 5.75 for SC pupils and 2.35 for non-SC pupils, close to the established-market estimates of 5.81 and 2.16. The non-SC zero-contact effect is larger, at 1.37 percentage points observed and 1.93 beyond random allocation. The broader sample also exhibits less stable pre-period behaviour, particularly for non-SC contact outcomes. I therefore use it as a complementary extension rather than allowing the larger extensive-margin estimates to determine the main specification.

The market-structure diagnostics point to a related difference. The differential pre-period

slope in the share of multi-school markets is 0.11 percentage points per year in established markets and 0.67 in the broader sample (Tables A10 and A11). Both slopes are positive, as are the corresponding pooled reform estimates, and both reject zero. Restricting to established markets reduces this movement; it does not eliminate it. The number of active schools also shows significant pre-period differences in both samples. These are diagnostics of local market structure, not the first-stage access variable: having multiple schools somewhere in a market is distinct from the share of residents with two public schools within one kilometre. They do not replace a sample-matched event study of access.

D.3 Pre-period behaviour, trend sensitivity, and coverage

Figure A10 reports annual reduced-form coefficients, relative to 2008, for the two samples. These are effects of baseline eligibility by calendar year, not annual IV effects per ten-percentage-point increase in access. Tables A10 and A11 report the joint tests of the 2006 and 2007 coefficients, the pre-period linear slopes, pooled and trend-adjusted reduced forms, and the trend-adjusted IV estimates.

For non-SC out-group schoolmates and beyond-random zero contact in established markets, the joint-test p-values are 0.788 and 0.711; the corresponding slope p-values are 0.529 and 0.516. These tests detect no differential pre-period movement in the two outcomes. The SC count and share do reject the joint test, with p-values of 0.0013 and 0.028. For the count, the pre-period pattern is non-monotone and its linear slope is not statistically significant. For the share, the pre-period slope is positive, at 0.120 percentage points per year ($p = 0.006$), whereas the pooled effect is negative. This change in direction is relevant to interpreting the event study. It does not establish that the earlier path would have continued without the reform or that the pooled estimate is a lower bound on the causal effect. Equally, a failure to reject a pretest is not evidence that every relevant counterfactual trend has been excluded.

The absence of statistical significance in the pooled non-SC share estimate should not be read as evidence of no composition response. Its conventional 95 percent interval is approximately $[-0.45, 0.05]$ percentage points. A decline of 1.4 percent of its baseline mean of 20.79 percent would be about 0.29 percentage points, which is inside that interval. The data therefore do not rule out a proportional decline comparable to the SC point estimate. This comparison is not a formal test of equality between the two group effects.

The trend specification adds baseline eligibility interacted with linear calendar time to the full pooled model, maintaining village and district–year fixed effects, controls, and each group's observations. It asks how much of the pooled response remains after allowing eligible and ineligible villages to follow different linear paths. Because the trend is estimated using post-reform as well as pre-reform years, it can absorb part of a gradually evolving policy response. It is not an extrapolation estimated solely from the pre-period or a definitive correction for confounding trends.

The sensitivity is most transparent in the reduced forms. In established markets, trend adjustment leaves the SC and non-SC out-group count effects at 30 and 23 percent of their pooled values. In contrast, the SC out-group-share effect remains at 96 percent, and the non-SC observed zero-contact effect at 111 percent. The latter two reduced-form magnitudes are

therefore much more stable than the counts. The non-SC beyond-random zero-contact effect retains 68 percent of its pooled value; it remains positive but is not unchanged. The SC zero-contact estimates retain about two-thirds to seven-tenths of their pooled magnitudes and are no longer statistically distinguishable from zero under conventional inference. School-size reduced forms become small and change sign, while own-group counts also change substantially. These patterns do not support a stable decomposition into school size and composition.

Reading only the IV column obscures this distinction. The estimated first-stage coefficient falls from approximately 0.33 to 0.18 in both groups. The IV estimate divides the reduced form by this access response, so a larger IV ratio can coexist with an unchanged or smaller eligibility effect. For example, the non-SC zero-contact reduced form is only modestly larger after adjustment, but dividing by a much smaller first stage roughly doubles the IV estimate. A larger ratio is not stronger evidence of a larger policy effect.

The established-market first-stage F statistic also falls from approximately 35.5 to 12.2. In the broader sample it falls to 8.4 for non-SC pupils and 10.2 for SC pupils. Reported standard errors and significance levels are conventional clustered inference, not Anderson–Rubin or other weak-instrument-robust confidence sets. The stars on the adjusted IV estimates should not be interpreted as results of such robust tests; no numerical correction to standard errors is imposed here. The reduced forms remain useful because they report the eligibility contrast without dividing by a weaker first stage. They still require the design's counterfactual-trends assumptions. In particular, stability of the SC share reduced form under a fitted trend does not resolve its pre-period differences.

Table A13 documents another limitation of the annual comparisons. The established-market core sample contains 335,672 villages in 2008, compared with 289,232 in 2006 and 255,187 in 2010. Its eligible share rises from 56.19 percent in 2008 to 58.37 percent in 2010. The broader sample also loses observations in 2010. These changes establish that the observed composition varies with year; they do not identify the direction of any resulting bias in the conditional estimates. I therefore do not use a pooled 2010–11 estimate as the main timing argument for contact. The full event-study paths are shown alongside the coverage audit, with the same caution about changing observations applying to pre-period tests and post-period coefficients.

D.4 Alternative definitions of low contact

Table A12 compares five definitions in the established-market sample: zero out-group pupils, fewer than three, fewer than five, and out-group enrolment shares at most 10 and 5 percent. For non-SC pupils, all five beyond-random effects are positive and statistically significant under the reported conventional inference. Their magnitudes range from 5.4 to 10.5 percent of their respective excess baseline means; the table reports each coefficient, standard error and denominator. This supports the non-SC low-contact finding across this specified set of definitions; it does not make the largest threshold coefficient the preferred estimate. The observed zero-contact effect of 0.66 percentage points remains the main extensive-margin illustration.

The SC results are less uniform. Share-based definitions have positive beyond-random estimates, but the estimates using fewer than three or fewer than five non-SC schoolmates are small or negative and statistically insignificant. Proportions and headcounts answer different

questions: a school can be heavily SC and still contain several non-SC pupils if it is sufficiently large. The difference between the two definitions does not, by itself, identify a change in school size or a mechanism of consolidation.

Figures A11 and A12 show the denser threshold grids estimated in the broader contact sample, with observed and excess outcomes presented separately. They are not event studies: each point is a pooled IV estimate for a different outcome definition. Non-SC excess effects are positive across the plotted grid, but not every observed headcount-threshold estimate is significant; for example, the effect at at most ten SC schoolmates has $p = 0.139$. The SC count curve is non-monotone. The intervals are pointwise, not simultaneous confidence bands for the full curve, and adjacent thresholds concern overlapping sets of pupils. I do not treat these estimates as independent replications, as proof of a distributional shift at every possible cutoff, or as a reason to select a favourable threshold.

Finally, the outcomes concern the schoolmates present in a given school year. The panel does not follow individual children through successive grades. Neither these estimates nor sums of grade-specific averages identify the probability that a particular child completes primary school without any out-group schoolmate. The contact interpretation remains one of school-level exposure, not accumulated individual interaction or downstream welfare.

Table A9: Cross-caste exposure in the broader contact sample

Outcome	2008 mean	RF (SE)	IV (SE)	IV / mean (%)
SC pupils: first stage 0.3136 (0.0578); $F = 29.46$				
Out-group schoolmates	92.01	-1.803^{***} (0.260)	-5.750^{***} (1.059)	-6.2
Out-group share (pp)	64.22	-0.406^{***} (0.081)	-1.294^{***} (0.225)	-2.0
Own-group schoolmates	44.25	0.680^{***} (0.157)	2.169^{***} (0.599)	4.9
School size	137.26	-1.123^{***} (0.349)	-3.581^{***} (1.153)	-2.6
Zero contact (pp)	2.60	0.108^{**} (0.051)	0.346^{**} (0.143)	13.3
Zero contact, excess (pp)	2.58	0.112^{**} (0.049)	0.357^{***} (0.133)	13.8
3,550,839 village–years; 382,630 villages; 552 district clusters.				
Non-SC pupils: first stage 0.2812 (0.0556); $F = 25.55$				
Out-group schoolmates	30.90	-0.661^{***} (0.118)	-2.351^{***} (0.467)	-7.6
Out-group share (pp)	21.53	-0.190^{***} (0.066)	-0.674^{***} (0.193)	-3.1
Own-group schoolmates	107.06	0.077 (0.321)	0.274 (1.141)	0.3
School size	138.96	−0.584 (0.355)	−2.077 (1.284)	-1.5
Zero contact (pp)	17.61	0.385^{***} (0.120)	1.370^{***} (0.329)	7.8
Zero contact, excess (pp)	15.16	0.544^{***} (0.108)	1.934^{***} (0.308)	12.8
3,700,508 village–years; 383,407 villages; 554 district clusters.				

Notes: The predetermined sample requires at least one active school in both 2006 and 2007 and both caste groups in the market in 2008. All outcomes within a group use a common complete-case sample. IV estimates are per ten-percentage-point increase in multiple-public-school access; RF estimates are effects of baseline eligibility. Models use the main controls and village and district–year fixed effects. Parentheses contain district-clustered standard errors. Stars use conventional inference: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. This broader population has less stable pre-period behaviour; these estimates are a complementary extension, not the main identification benchmark.

Table A10: Pre-period diagnostics and trend sensitivity: established markets

Outcome	Joint pre-test p	Pre-period slope (p)	Pooled RF (SE)	Trend RF (SE)	Trend IV (SE)
Market structure (core sample): pooled $F = 35.47$; trend $F = 12.18$					
		−0.075	0.037*	−0.042**	−0.234*
Number of active schools	< 0.001	($p < 0.001$)	(0.021)	(0.018)	(0.123)
		0.111	0.647***	0.229***	1.286***
Multi-school market share (pp)	< 0.001	($p < 0.001$)	(0.085)	(0.054)	(0.460)
SC pupils: pooled $F = 35.50$; trend $F = 12.19$					
		0.250	−1.915***	−0.573**	−3.207**
Out-group schoolmates	0.001	($p = 0.144$)	(0.249)	(0.242)	(1.361)
		0.120	−0.301***	−0.288***	−1.615***
Out-group share (pp)	0.028	($p = 0.006$)	(0.072)	(0.078)	(0.481)
		−0.036	0.254*	0.587***	3.287***
Own-group schoolmates	0.108	($p = 0.781$)	(0.154)	(0.142)	(1.188)
		0.214	−1.661***	0.014	0.080
School size	< 0.001	($p = 0.374$)	(0.337)	(0.289)	(1.623)
		−0.017	0.095**	0.068	0.379
Zero contact above random (pp)	0.125	($p = 0.591$)	(0.048)	(0.055)	(0.304)
		−0.021	0.092*	0.062	0.345
Zero out-group schoolmates (pp)	0.121	($p = 0.510$)	(0.049)	(0.054)	(0.297)
Non-SC pupils: pooled $F = 35.49$; trend $F = 12.18$					
		−0.041	−0.710***	−0.163*	−0.913*
Out-group schoolmates	0.788	($p = 0.529$)	(0.112)	(0.091)	(0.481)
		0.003	−0.065	−0.084*	−0.473*
Out-group share (pp)	0.056	($p = 0.896$)	(0.045)	(0.046)	(0.256)
		0.193	−0.638**	0.331	1.857
Own-group schoolmates	< 0.001	($p = 0.376$)	(0.282)	(0.258)	(1.602)
		0.151	−1.348***	0.168	0.945
School size	0.003	($p = 0.539$)	(0.331)	(0.284)	(1.677)
		−0.035	0.293***	0.199*	1.114**
Zero contact above random (pp)	0.711	($p = 0.516$)	(0.081)	(0.110)	(0.565)
		−0.101	0.217**	0.240**	1.347**
Zero out-group schoolmates (pp)	0.182	($p = 0.059$)	(0.085)	(0.108)	(0.587)

Notes: RF columns report eligibility effects; the IV column scales by the trend-adjusted first stage and is per ten-percentage-point increase in access. The joint test concerns the 2006 and 2007 event-study coefficients relative to 2008. Slopes are pre-period linear-slope diagnostics in outcome units per year. The trend specification adds baseline eligibility interacted with calendar time to the full panel, not a pre-period-only extrapolation. Comparisons retain a common sample within each panel; the core market sample differs from the group-specific contact samples. Market structure is not the resident-level public-school access measure used in the first stage. Parentheses contain district-clustered standard errors. Stars use conventional inference ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$), not weak-instrument-robust tests. A failure to reject a pretest does not establish parallel trends.

Table A11: Pre-period diagnostics and trend sensitivity: broader contact sample

Outcome	Joint pre-test p	Pre-period slope (p)	Pooled RF (SE)	Trend RF (SE)	Trend IV (SE)
Market structure (core sample): pooled $F = 25.44$; trend $F = 8.39$					
Number of active schools	< 0.001	-0.084 ($p < 0.001$)	0.060^{***} (0.022)	-0.074^{***} (0.020)	-0.491^{**} (0.220)
Multi-school market share (pp)	< 0.001	0.673 ($p < 0.001$)	4.169^{***} (0.312)	2.305^{***} (0.313)	15.252^{***} (3.826)
SC pupils: pooled $F = 29.46$; trend $F = 10.24$					
Out-group schoolmates	0.001	0.338 ($p = 0.052$)	-1.803^{***} (0.260)	-0.540^{**} (0.264)	-3.136^{**} (1.480)
Out-group share (pp)	0.024	0.114 ($p = 0.009$)	-0.406^{***} (0.081)	-0.355^{***} (0.086)	-2.061^{***} (0.565)
Own-group schoolmates	0.078	0.022 ($p = 0.866$)	0.680^{***} (0.157)	0.707^{***} (0.141)	4.105^{***} (1.375)
School size	< 0.001	0.360 ($p = 0.142$)	-1.123^{***} (0.349)	0.167 (0.297)	0.969 (1.819)
Zero contact above random (pp)	0.123	-0.010 ($p = 0.724$)	0.112^{**} (0.049)	0.134^{**} (0.053)	0.779^{**} (0.347)
Zero out-group schoolmates (pp)	0.034	-0.038 ($p = 0.209$)	0.108^{**} (0.051)	0.139^{**} (0.054)	0.806^{**} (0.360)
Non-SC pupils: pooled $F = 25.55$; trend $F = 8.40$					
Out-group schoolmates	0.001	0.271 ($p < 0.001$)	-0.661^{***} (0.118)	-0.259^{**} (0.109)	-1.711^{**} (0.716)
Out-group share (pp)	< 0.001	0.105 ($p = 0.001$)	-0.190^{***} (0.066)	-0.254^{***} (0.063)	-1.683^{***} (0.534)
Own-group schoolmates	0.077	0.198 ($p = 0.363$)	0.077 (0.321)	0.688^{**} (0.267)	4.553^{*} (2.392)
School size	0.002	0.469 ($p = 0.052$)	-0.584 (0.355)	0.429 (0.291)	2.842 (2.308)
Zero contact above random (pp)	0.038	-0.159 ($p = 0.011$)	0.544^{***} (0.108)	0.430^{***} (0.130)	2.848^{***} (0.899)
Zero out-group schoolmates (pp)	< 0.001	-0.615 ($p < 0.001$)	0.385^{***} (0.120)	0.472^{***} (0.130)	3.125^{***} (1.030)

Notes: RF columns report eligibility effects; the IV column scales by the trend-adjusted first stage and is per ten-percentage-point increase in access. The joint test concerns the 2006 and 2007 event-study coefficients relative to 2008. Slopes are pre-period linear-slope diagnostics in outcome units per year. The trend specification adds baseline eligibility interacted with calendar time to the full panel, not a pre-period-only extrapolation. Comparisons retain a common sample within each panel; the core market sample differs from the group-specific contact samples. Market structure is not the resident-level public-school access measure used in the first stage. Parentheses contain district-clustered standard errors. Stars use conventional inference ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$), not weak-instrument-robust tests. A failure to reject a pretest does not establish parallel trends.

Table A12: Alternative low-contact definitions in established markets

Definition	Observed mean	Observed IV (SE)	Excess mean	Excess IV (SE)	Excess IV / mean (%)
SC pupils					
Zero out-group pupils	2.85	0.278** (0.136)	2.83	0.288** (0.131)	10.2
Fewer than 3 out-group pupils	3.83	−0.091 (0.206)	3.71	0.018 (0.171)	0.5
Fewer than 5 out-group pupils	4.79	−0.460 (0.280)	4.47	−0.219 (0.219)	-4.9
At most 10% out-group enrolment	5.39	0.346* (0.187)	5.12	0.487*** (0.173)	9.5
At most 5% out-group enrolment	4.02	0.191 (0.166)	3.92	0.314** (0.147)	8.0
Non-SC pupils					
Zero out-group pupils	19.24	0.658*** (0.215)	16.58	0.892*** (0.203)	5.4
Fewer than 3 out-group pupils	24.96	0.698*** (0.244)	16.13	1.245*** (0.251)	7.7
Fewer than 5 out-group pupils	29.69	1.015*** (0.268)	15.44	1.625*** (0.290)	10.5
At most 10% out-group enrolment	39.55	1.149*** (0.291)	16.58	1.463*** (0.301)	8.8
At most 5% out-group enrolment	29.46	1.136*** (0.250)	17.18	1.441*** (0.273)	8.4

Notes: Means refer to 2008. Observed and excess means and coefficients are in percentage points; IV effects are per ten-percentage-point increase in access. The last column divides the excess IV coefficient by the corresponding excess mean, not by the observed rate. Excess subtracts the exact random-allocation expectation conditional on market caste totals and school enrolments. Fewer than three and fewer than five mean at most two and at most four, respectively. Share thresholds include the boundary. Each group uses the same sample as its panel in Table 4. Stars and standard errors follow the main table and are pointwise; no multiple-testing adjustment is imposed. The reported grid is a sensitivity analysis, not a justification for selecting a cutoff with a large estimate.

Table A13: Coverage in selected pre- and early post-reform years

Sample	Year	Villages	District clusters	Eligible share (%)
Broader contact	2006	383,407	554	58.92
Broader contact	2007	383,407	554	58.92
Broader contact	2008	383,407	554	58.92
Broader contact	2010	301,222	505	61.08
Broader contact	2011	383,295	554	58.91
Established markets	2006	289,232	508	54.65
Established markets	2007	331,145	536	56.15
Established markets	2008	335,672	542	56.19
Established markets	2010	255,187	485	58.37
Established markets	2011	316,631	523	55.30

Notes: Figures describe the core estimation sample used for the pre-period diagnostics, not either group's outcome-specific complete-case sample. Coverage differs across years despite predetermined membership rules. Changes in eligible shares document sample composition; they do not by themselves identify the direction or magnitude of selection bias.

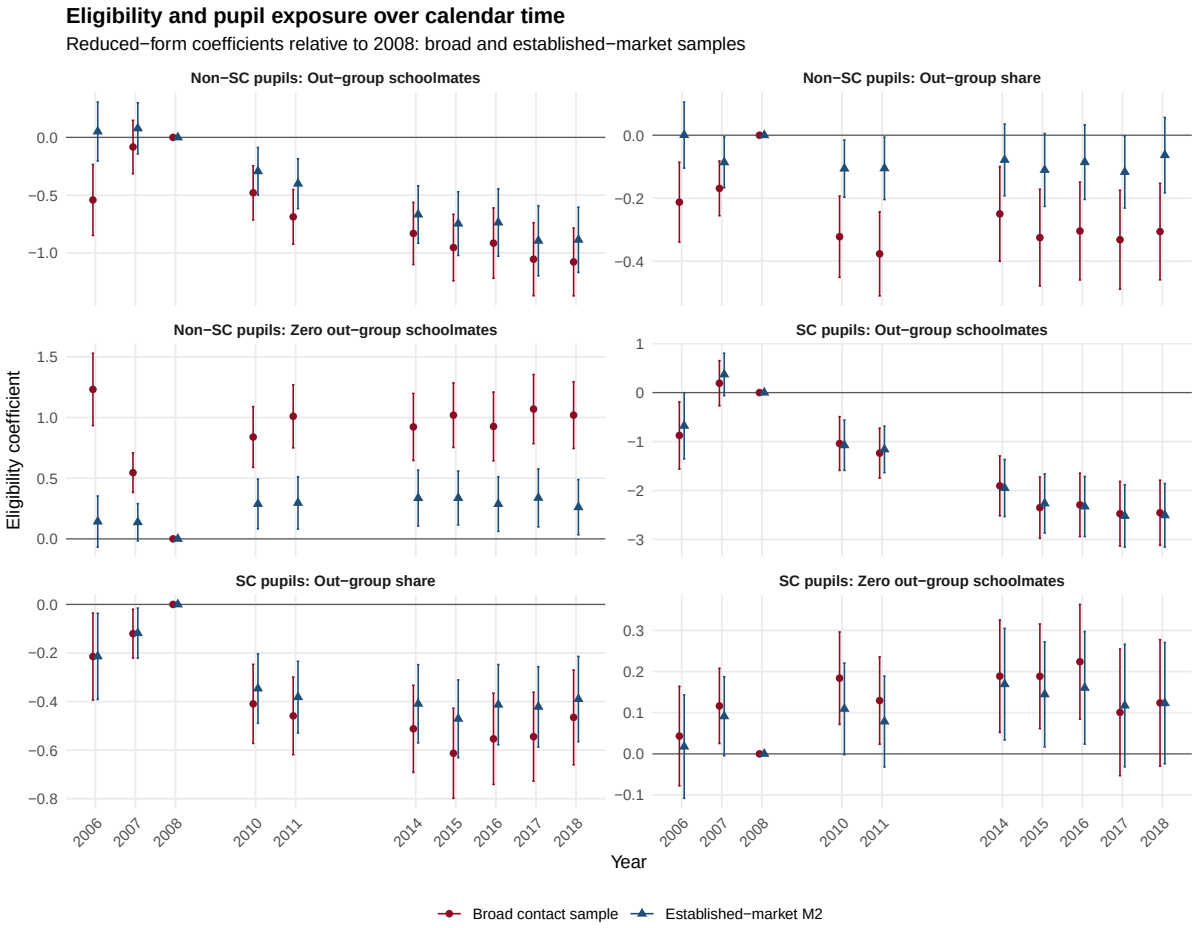


Figure A10: Annual eligibility effects on contact: broader and established-market samples

Notes: The supplied event-study figure reports reduced forms relative to 2008. Coefficients are not annual IV ratios. Outcomes are measured at the village–year level using within-market pupil weights. Intervals are pointwise; coverage varies across years as documented in Table A13.

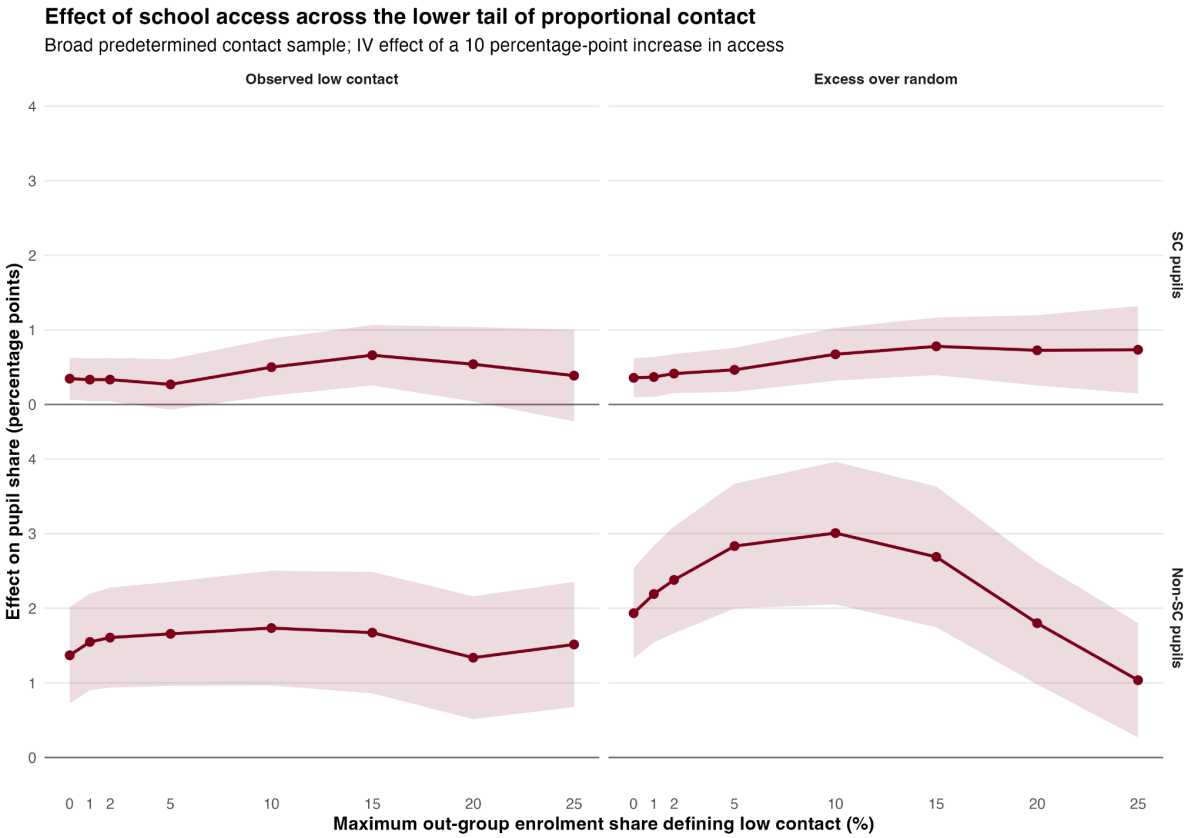


Figure A11: Share-based low-contact thresholds in the broader contact sample

Notes: Each point is a pooled IV estimate per ten-percentage-point access increase, not an annual effect. Thresholds are maximum out-group enrolment shares. The sample is broader than the established-market sample used in the main table. Shaded intervals are pointwise 95 percent confidence intervals using conventional clustered inference, with no simultaneous-band or multiple-testing adjustment.

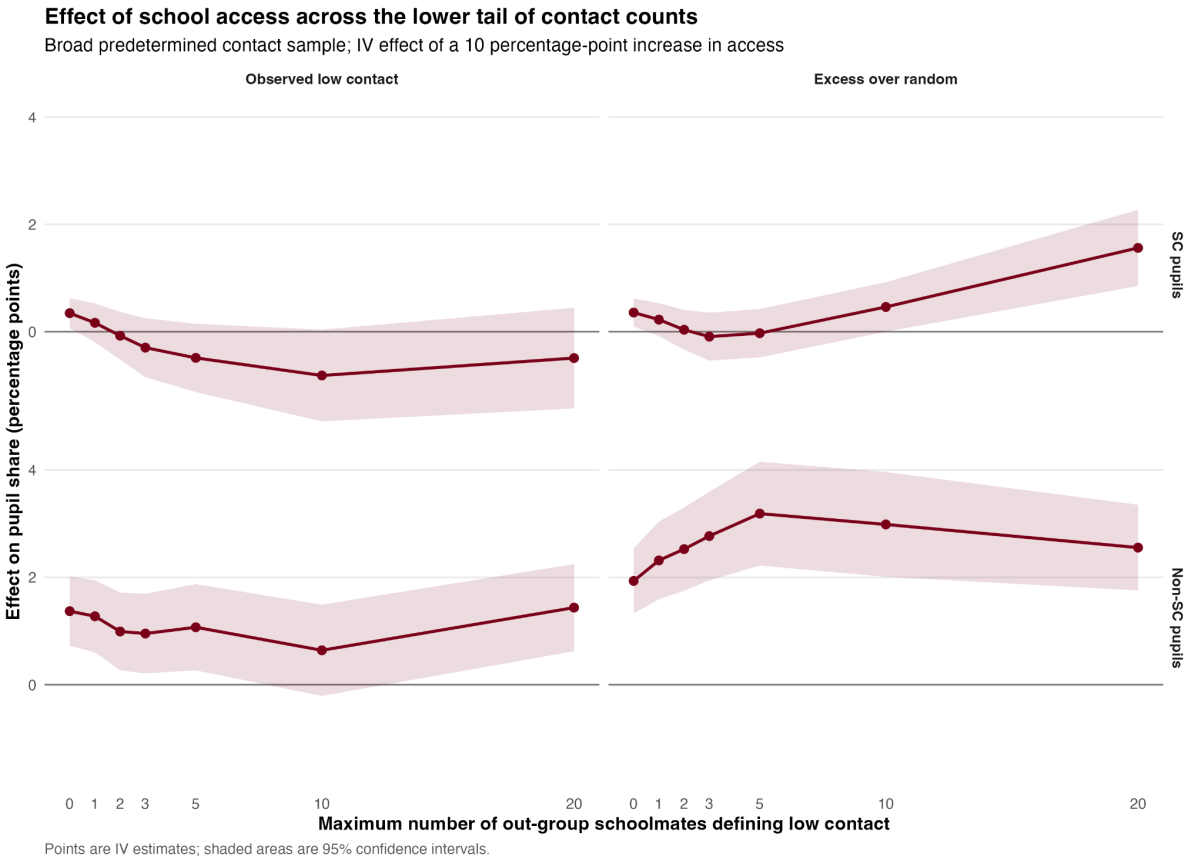


Figure A12: Headcount-based low-contact thresholds in the broader contact sample

Notes: Thresholds are maximum numbers of out-group schoolmates. Zero is exactly zero, and a threshold of two corresponds to fewer than three. The estimates are pooled IV effects, on the broader contact sample, with pointwise 95 percent intervals; see the notes to Figure A11.

E Residential Geography and Geographic Assignment

This section provides the construction details for the geographic-assignment evidence in Section 6. The exercises distinguish the location of the geographically closest school from observed enrolment. They therefore do not impose attendance boundaries or claim that households necessarily enrol at the nearest school.

E.1 Direct localization of the public-school network

The direct-localization exercise uses inhabited 100-metre cells from the Global Human Settlement Layer in 2008 and 2018. For each cell, I identify the nearest active public primary school in the corresponding year and calculate the straight-line distance to it. I then overlay both the cell centroid and the school coordinate on the Census-village boundary layer. Comparing the two village identifiers indicates whether the nearest public school is located in the resident's own Census village or in another village. Cells with zero estimated population are excluded.

I aggregate the cell-level measures to the Census-village–year level using GHSL population weights. This produces the population-weighted mean distance to the nearest public primary school and the population shares whose nearest public school is located inside and outside the resident's Census village. The inside- and outside-village shares are complements among cells for which both village identifiers are observed. The reduced-form regressions use the same fixed effects, village-characteristic-by-year controls, and district clustering as the headline specification. Population weights enter the construction of each village-level outcome; the regressions do not weight villages by their total population. Table A14 reports the resulting estimates.

The school universe is deliberately restricted to public primary schools. The policy directly changed the density and location of this network, whereas private entry and exit may themselves respond to the reform. The exercise thus tests whether the public network altered by the mandate became more local. It does not establish that the entire public–private market became more local or that cross-village enrolment declined.

E.2 Coarse and habitation-calibrated residential benchmarks

The coarse benchmark assigns every inhabited cell the SC share of its Census village. It therefore preserves the observed difference in caste composition across Census villages while imposing complete caste mixing within each village. Residents are assigned to the nearest active public primary school, school caste composition is reconstructed from the assigned cell populations, and SC–non-SC dissimilarity is calculated over the resulting school market. Residential population and caste composition are fixed. Across years, only the set and locations of active schools change as schools open. The simulated outcome is estimated using the same reduced-form specification and the same first-stage-locked sample as observed dissimilarity.

The fine benchmark uses the Jal Jeevan Mission habitation directory to discipline caste differentiation below the Census-village level. The directory records SC, ST, and general-category populations for approximately 1.67 million named rural habitations but does not provide household coordinates. Habitation locations are approximated with geotagged non-school facilities carrying habitation names. Anganwadi coordinates are preferred and water-supply points are

Table A14: Direct localization of the public-school network

	Eligibility × Post
Nearest public school outside Census village	-0.0381^{***}
	(0.0034)
Nearest public school inside Census village	0.0381^{***}
	(0.0034)
Distance to nearest public school (metres)	-64.6^{***}
	(5.4)

Notes: Each row reports a separate reduced-form regression using the 2008 and 2018 endpoints. The first two outcomes are population-weighted shares of inhabited 100-metre cells whose nearest active public primary school lies outside or inside the resident's Census village. These coefficients are reported as proportions, so 0.0381 corresponds to 3.81 percentage points. The third outcome is the population-weighted mean distance to the nearest active public primary school, measured in metres. Population weights enter the construction of the village-level outcomes; the regressions retain the village–year as the unit of observation and do not weight villages by total population. All specifications include village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls. Standard errors are clustered by district. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

used as a fallback. Names are matched within the administrative hierarchy; invalid coordinates are discarded, and labels whose facility points are dispersed by more than five kilometres are rejected. The matching and calibration do not use school coordinates, treatment assignment, enrolment, or school-segregation outcomes.

For a Census village v with habitations h, strict habitation dissimilarity is

$$D_v^{JJM} = \frac{1}{2} \sum_{h \in v} \left| \frac{SC_{hv}}{SC_v} - \frac{NSC_{hv}}{NSC_v} \right|. \tag{26}$$

The calibrated rule assigns a radial caste gradient to the inhabited GHSL cells while preserving each Census village's aggregate SC share exactly. The parameter λ controls the strength of within-village differentiation; $\lambda = 0$ reproduces the coarse benchmark. The headline value is selected by

$$\lambda^* = \arg\min_{\lambda} \left| \overline{D^{sim}(\lambda)} - \overline{D^{JJM}} \right|, \tag{27}$$

using the 23,225 Census villages that satisfy the calibration requirements. Mean strict JJM dissimilarity is 0.245. The simulated means are 0.210, 0.241, and 0.271 at $\lambda = 0.70$, 0.80, and 0.90, respectively, selecting $\lambda = 0.80$. District bootstrap resampling places the central calibration range at approximately 0.70–0.90. Only after this selection is fixed is the residential rule transported to spatially usable villages in the main estimation sample and combined with the realised school network.

The calibration does not claim to recover the true location of caste groups in every village. Facility coordinates are proxies for habitation locations, and the radial relationship estimated in the matched data is transported to unmatched villages. Its role is narrower: it provides an external, school-blind discipline for the degree of within-village caste differentiation used in the fine benchmark. Table A15 reports the resulting estimates across the calibration range.

E.3 School-universe robustness

The headline assignment benchmark uses the nearest active public primary school because the mandate directly expanded public provision. As a robustness check, I repeat the assignment after allowing the closest active school to be either public or private. This changes neither the residential input nor the calibrated value of λ; it changes only the set of schools available to the common distance rule. The public-plus-private variants are reported in Table A16. They align the assignment universe with the full public–private market used to construct observed dissimilarity while retaining the public-only exercise as the direct test of localization in the network affected by the mandate.

Table A15: Geographic-assignment benchmark: sensitivity to within-village caste differentiation

	Observed	Nearest-public-school benchmark			
	Segregation	$\lambda = 0$	$\lambda = 0.70$	$\lambda = 0.80$	$\lambda = 0.90$
Reduced-form effect	0.00550***	0.00470***	0.00524***	0.00541***	0.00559***
	(0.00082)	(0.00055)	(0.00059)	(0.00060)	(0.00061)
Share of observed response	100.0%	85.5%	95.4%	98.5%	101.7%
Residual: observed minus benchmark	–	0.00080	0.00026	0.00008	-0.00009
		(0.00065)	(0.00066)	(0.00067)	(0.00067)
Villages	347,829	347,829	347,829	347,829	347,829
District clusters	540	540	540	540	540
Observations	3,169,565	3,169,565	3,169,565	3,169,565	3,169,565

Notes: The table reports reduced-form effects of baseline eligibility interacted with the post-reform period on observed or geographically assigned SC–non-SC dissimilarity. The observed column reports realised school segregation. Benchmark columns assign every inhabited 100-metre cell to its nearest active public primary school. At $\lambda = 0$, each cell inherits its Census village's SC share. Positive values of λ introduce within-village caste differentiation while preserving each village's aggregate SC share. The habitation-calibrated headline value is $\lambda = 0.80$. The residual is observed dissimilarity minus the corresponding benchmark outcome, estimated directly using the same specification. The share of the observed response is the ratio of the benchmark and observed reduced-form coefficients; it is not an R^2 or a causal mediation share. All specifications use village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls, exclude 2009, and cluster standard errors by district. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A16: Geographic-assignment benchmark: sensitivity to the school universe

	Observed	Geographic-assignment benchmark	
	Segregation	Public schools	Public and private schools
Reduced-form effect	0.00550***	0.00542***	0.00553***
	(0.00082)	(0.00060)	(0.00058)
Share of observed response	100.0%	98.6%	100.5%
Residual: observed minus benchmark	–	0.00008	-0.00003
		(0.00067)	(0.00066)
Villages	347,771	347,771	347,771
District clusters	540	540	540
Observations	3,169,163	3,169,163	3,169,163

Notes: The table holds the within-village caste-differentiation parameter at the habitation-calibrated value, $\lambda = 0.80$, and varies the set of schools used for geographic assignment. The public-schools column assigns every inhabited 100-metre cell to its nearest active public primary school. The public-and-private-schools column instead assigns each cell to its nearest active primary school in either sector. The observed column reports realised school segregation. The residual is observed dissimilarity minus the corresponding benchmark outcome, estimated directly using the same specification. The share of the observed response is the ratio of the benchmark and observed reduced-form coefficients; it is not an R^2 or a causal mediation share. All specifications use village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls, exclude 2009, and cluster standard errors by district. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

F Alternative Mechanisms and School-Input Diagnostics

The main text uses a fixed pre-reform observable-input index to ask whether the reform shifted SC and non-SC pupils towards incumbent public schools that were already observably better or worse. This section reports complementary diagnostics. It first tests whether the segregation coefficient is sensitive to conditioning on contemporaneous inputs. It then reports changes in the individual input measures and verifies that coverage of the predetermined exposure measure is balanced across caste groups.

F.1 Controlling for contemporaneous observable inputs

Table A17 compares the baseline segregation estimate on a common input-observed sample with a specification that adds the contemporaneous observable-input index. The estimate changes little. Because contemporaneous inputs may themselves respond to the reform, the controlled specification is a mechanism diagnostic rather than the preferred causal specification.

F.2 Individual observable inputs in incumbent public schools

The aggregate input index can mask offsetting changes across components. Table A18 therefore reports each observable input separately. The pattern is mixed rather than uniformly favourable or unfavourable. These are school-side responses among incumbents and should not be interpreted as predetermined quality differences selected by families.

Table A17: Robustness of the segregation estimate to observable school inputs

	Baseline (1)	+ Observable-input index (2)
Eligibility × Post first stage	0.3282***	0.3266***
	(0.0550)	(0.0547)
First-stage F statistic	35.6	35.6
2SLS: observed dissimilarity	0.0164***	0.0161***
	(0.0022)	(0.0022)
Mean observed dissimilarity	0.430	0.430
Villages	349,180	349,180
District clusters	550	550
Observations	3,128,350	3,128,350

Notes: The unit of observation is a village–year in the predetermined established-market sample; 2009 is excluded. The dependent variable is observed SC–non-SC dissimilarity. Both columns use the same sample. Column 1 reports the baseline specification. Column 2 adds the contemporaneous observable-input index for incumbent public primary schools within one kilometre. Multiple-school access is expressed in units of 10 percentage points and instrumented with baseline eligibility interacted with the post-reform period. Both specifications include village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls. Standard errors are clustered by district. Because the contemporaneous input index may respond to treatment, Column 2 is a mechanism diagnostic rather than the preferred causal specification. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A18: Effects on observable inputs in incumbent public schools

	First stage	Reduced form	2SLS	Mean	F	Villages	Clusters	Observations
Composite input index	0.3282***	-0.0035***	-0.0106***	0.618	35.6	349,180	550	3,128,350
	(0.0550)	(0.0007)	(0.0019)					
Usable classrooms	0.3282***	0.0027**	0.0082**	0.713	35.6	349,180	550	3,128,417
	(0.0550)	(0.0012)	(0.0037)					
Adequate boundary wall	0.3288***	0.0001	0.0003	0.368	35.8	349,258	550	3,130,618
	(0.0549)	(0.0013)	(0.0041)					
Adequate drinking water	0.3286***	-0.0078***	-0.0237***	0.832	35.8	349,261	550	3,132,162
	(0.0549)	(0.0011)	(0.0050)					
Library	0.3416***	-0.0088***	-0.0257***	0.724	37.5	349,241	550	3,056,291
	(0.0558)	(0.0018)	(0.0053)					
Playground	0.3286***	-0.0040***	-0.0120***	0.563	35.8	349,261	550	3,132,006
	(0.0549)	(0.0014)	(0.0045)					
Ramps	0.3288***	-0.0028*	-0.0084*	0.623	35.8	349,256	550	3,129,715
	(0.0549)	(0.0016)	(0.0044)					
Books intensity	0.3282***	-0.0065***	-0.0198***	0.488	35.6	349,180	550	3,128,419
	(0.0550)	(0.0012)	(0.0035)					

Notes: Each row reports a separate village–year regression for the indicated contemporaneous observable input among incumbent public primary schools in the one-kilometre market. The composite index is formed from usable classrooms, adequate boundary wall, adequate drinking water, library, playground, ramps, and books intensity; it is defined when at least six of the seven components are available. The first-stage column reports the effect of baseline eligibility interacted with the post-reform period on access to at least two nearby public primary schools, expressed in units of 10 percentage points. The reduced-form column reports the effect of the same interaction on the input outcome. The 2SLS column instruments multiple-school access with the eligibility interaction. Each row uses its own outcome-available sample. All specifications use the established-market sample, exclude 2009, and include village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls. Standard errors are clustered by district. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

F.3 Coverage of the predetermined input-exposure measure

The fixed pre-reform input-exposure analysis requires an observable-input index for incumbent schools. Table A19 tests whether coverage differs systematically between SC and non-SC enrolment. The outcome is the non-SC coverage share minus the SC coverage share. Its mean is close to zero, and the reform does not create a relative loss of coverage for SC pupils. The caste-specific exposure result is therefore not generated by systematically worse measurement of the schools attended by SC pupils.

Table A19: Coverage of the caste-specific observable-input exposure measure

	Non-SC − SC coverage gap
Eligibility × Post first stage	0.3281***
	(0.0549)
First-stage F statistic	35.7
Eligibility × Post reduced form	-0.0015**
	(0.0006)
Mean of outcome	-0.003

Notes: The outcome is the share of non-SC enrolment minus the share of SC enrolment in incumbent public primary schools for which the fixed 2006–2008 observable-input index is available. A negative outcome indicates better coverage of SC enrolment. The reduced-form coefficient reports the effect of baseline eligibility interacted with the post-reform period. The first-stage row reports the corresponding effect on multiple-school access, expressed in units of 10 percentage points. The specifications use the established-market sample, exclude 2009, and include village fixed effects, district-by-year fixed effects, and village-characteristic-by-year controls. Standard errors are clustered by district. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.