Online Appendix

1 Construction of the Dataset

1.1 Data Sources

We constructed the block-level dataset from the 1991 and 2001 Censuses of India ("Population Census"); the 2002-03, 2007-08, and 2008-09 rounds of the District Information System for Education (DISE or "School Census"); and the NPEGEL and KGBV lists for 2007-08 from the Ministry of Human Resource Development (MHRD).

The censuses contained the literacy rates, data to estimate the school age population in the two periods, and other population covariates. The data CDs were provided by the data dissemination unit of the Ministry of Home Affairs, Office of the Registrar General, in New Delhi. The Population Census of 1991 included block-level data, but public availability of the 2001 data below the district division was very restricted. Only a limited set of variables included in the "Primary Census Abstract" was released at the village level, but without block location codes. Country-wide data at the block level was not distributed at all. We located the block codes in another census source ("Village and Town Location Codes" CD), which allowed us to calculate EBB status and other block values from village-level data.

The DISE was the primary source for education outcomes, program resources, and other school characteristics. This School Census was launched by the National University of Educational Planning and Administration (NUEPA) with support of the MHRD. Each year, headmasters filled in the questionnaires for the reference date September 30. Data entry largely took place in the districts through a standardized application (http://www.dise.in). Completeness and consistency checks were carried out at all stages up to the national level, and external agencies randomly verified 5 percent of the sample. The national data server is located at NUEPA in New Delhi, from where we got a copy of the complete DISE raw data until 2008-09. This was necessary since the website (http://www.schoolreportcards.in) permitted access to raw data only for a limited set of variables and from 2005-06 onwards. Pre-program rounds prior to 2002-03 did not have sufficient coverage, and post-program years before 2007-08 would not have added identifying variation due to missing program participation status.

Program documentation, details, and data were provided by the national gender consultant of the SSA, Kiran Dogra, and other contacts in the Ministry. Some of the information has now also been published on the SSA website (http://ssa.nic.in). NPEGEL/KGBV participation denotes whether program activities had been approved by 2007-08 and is not available for earlier years. The government records did not identify which NPEGEL blocks were not yet operational, but indicated that these units accounted for less than 1 percent of approved blocks. More information at the block level is not centrally available, but only in the about 600 district offices of India. Besides approval status, the data tables in MHRD (2008e) essentially included details on SSA, NPEGEL, and KGBV financing and implementation aggregated by federal state.

1.2 Matching of the Administrative Division

To construct the dataset, we first linked blocks by location identifiers across data sources and then generated block-level values where not directly available. The upper part of Figure A1 illustrates how the full support of 4,001 blocks for the analysis was created from the original list of 6,367 blocks in the 2001 Population Census. We used the 2007-08 DISE, the most complete data source, as an anchor for the other data tables. The program had never been implemented in 10 smaller federal states with 251 blocks, which we hence excluded. The status of KGBV and NPEGEL approval for 2007-08 was available for the entire program area, except for the NPEGEL in two districts with 35 blocks. This program data was added to the DISE by manual string matching on state, district, and block names.

The key challenge was to link the administrative division in the 2007-08 DISE with the 2001 Population Census, given that correspondence tables do not exist. The two sources use the same location codes only up to the district level, requiring us to match the average of 10 blocks within districts by name. Discrepancies at this stage were the main cause of sample size reductions. They occurred largely because official block and district boundaries and/or names had changed over time, most commonly when new blocks had been carved out ("bifurcated") from two or more existing blocks. After a round of automatic string matching, we manually connected matches that were previously unsuccessful due to evident spelling discrepancies and some name changes verified in Google. In addition, we also traced blocks that had been entirely transferred to neighboring districts.

In the merge process it turned out that the block and district division in 4 federal states of the Himalaya region, and one in the South, had shifted so much over the study period that blocks were not traceable. These states, with 873 census blocks, were hence not included.[1]

In the remaining program states, we were able to match about 99 percent of the blocks with certainty. The few unassigned blocks left over from this procedure in 2007-08 were most likely bifurcated parts, which may have created some imputation errors in outcomes for their unidentified blocks of origin. Within the remaining states, another 26 blocks had either not been covered in that School

[1]For the southern state of Tamil Nadu, changes in the divison initially seemed somewhat less dramatic. After an attempt at generating block-level values, the resulting outcomes showed a fair amount of large outliers and suspiciously large means and standard deviations in comparison with the other 14 states, most likely from too many undetected boundary changes and some misreporting.

Census year (5 blocks) or spanned over two districts (21 blocks). Since these 21 pairs appear as two blocks each in the Population Census, but as one unit in the DISE, we recalculated the Population Census variables for the unit and considered it a program participant if at least one of its two parts was. The School Census, unlike the rural division in the Population Census, also covered urban blocks that we identified by their name affixes and their share of urban schools (usually > 90 percent) to exclude them.

After dropping states not in the 2002-03 DISE (see below), we linked the Population Census of 1991 through the 2001 round with the dataset. Since location codes were not identical in the two, we again matched on names within states. 212 blocks from the 2001 Population Census could not be identified in 1991.

Finally, we linked the School Census over time by location codes and checked their names for consistency. In addition to the previous issues, some areas were still missing in 2002-03 because the database was not officially complete until 2005-06. 6 smaller states with 272 blocks had not yet been covered in 2002-03, which finally left 14 large federal states (83.4 percent of the rural population) in the sample area.[2] Even within this area, 328 blocks were not yet included in the pre-intervention DISE although they appeared in the other data sources. Another 343 observations are incomplete for other reasons, and 26 are evident outliers in outcomes (9 in the estimation sample with $h = 8$). Adding the 2008-09 round for the few variables reported with 1 year lag did not further reduce the estimation sample (not shown in the graph). The complete matching procedure yielded a final dataset with 4,001 blocks.

[2] The states of Chhattisgarh, Jharkhand, and Uttarakhand were created from existing states only in November 2000. Since they continued to share similar educational institutions we defined state fixed effects based on the division prior to that month.

1.3 Generation of Block-Level Values

The Population Census of 1991 and the program data contained block-level values, but the Population Census of 2001 and the School Census needed to be aggregated from individual villages and schools, respectively.

Since the census office had only shared the literacy rates of the EBBs with the Ministry - but not of the non-EBBs - we first computed these (and other) variables for all rural blocks in India from the village data and checked consistency of the EBB classification with the ministry list.

For the School Census, we comprehensively checked the coding of categorical variables and other consistency issues. Where necessary for a given variable, we excluded a small fraction of inconsistent school observations before aggregating the rest into block values. The data cleaning did not affect the enrollment variables, except for recovering 1 percent of upper primary schools initially not classified as such, but reporting non-zero enrollment in these grades.

Block values were calculated from schools in the NPEGEL/KGBV target group, as shown in Figure A1: public centers in rural areas for upper primary girls. Consequently, we first discarded units in smaller urban settlements that the DISE (not the Population Census) sometimes covered within rural blocks. We selected only upper primary schools and upper primary children from schools with both sections. The DISE included KGBV schools only optionally, but made them identifiable by their classification and/or school name. By 2007-08, only about half of the operating KGBV units had submitted DISE data. We dropped these data and included instead complete girls' enrollment for all KGBV centers from the Ministry.[3]

[3] In the EBB area of the broadest band ($h = 8$ percentage points), KGBV enrollment is 0.93 percent of the female population of upper primary age. Using only partial KGBV enrollment from the School Census, rather than from the Ministry, would underestimate the program effect on enrollment.

2 Cost-Effectiveness

This section provides details on the calculation of cost-effectiveness estimates shown in Table 8.

2.1 Meller and Litschig (2014)

From April 2004 to March 2007, the central government spent 14.1 billion Rupees on the NPEGEL and KGBV programs (MHRD 2007, 2008d, 2008e). Federal states were required to spend an additional 4.7 billion Rupees to pay for teacher wages and maintenance. Using the nominal exchange rate corresponding to each year of outlay (45 Rupees per U.S. dollar in each year), and adjusting for U.S. inflation using the GDP deflator, total program expenditure over this three-year period was 427 million 2007 U.S. dollars. With an estimated 14 million girls in the target population, the annual cost per girl was about \$10. Since there is some evidence that the program also benefited boys, we compute separate cost-effectiveness ratios for girls and boys. Under the assumptions that there were roughly as many boys as girls in the target population and that 30% of NPEGEL/KGBV funds were spent on boys and 70% on girls, the annual program cost per boy was about 3 U.S. dollars and about 7 U.S. dollars for girls.

While most of the NPEGEL/KGBV funds were used to pay for current costs, some expenditures were also made for the KGBV school construction and NPEGEL infrastructure upgrading for existing schools. From the program guidelines (MHRD 2008a, 2008b), the one-time setup costs of these items corresponded to approximately 65% of the annual recurring costs of other program resources, for both the NPEGEL and KGBV. The share of setup costs in total program spending over the 3-year fiscal period was therefore $0.65/(0.65 + 3) = 0.18$. The setup cost per girl and year was $\$7 \times 0.18 = \$1.26$ and $\$3 \times 0.18 = \$0.54$ for boys. Using straight-line depreciation and assuming that the capital expenditures provide services for 30 years, the annual expense per girl was $\$1.26 \times 3/30 = \$0.126$ and for boys $\$0.54 \times 3/30 = \$0.054$. Together with the annual running cost per girl of $\$7 \times 0.82 = \$5.74$, the adjusted annual program cost per girl was about 5.87 U.S. dollars. For boys, the adjusted annual program cost was $\$3 \times 0.82 + \$0.054 = \$2.51$.

Given an estimated net enrollment gain for girls of about 4.5 percentage points from Table 6, the implied cost of enrolling an additional girl in upper primary school was about $\$5.87/0.045 = \$130$. If enrollment translates into grade completion one for one, we can also compute the years of girls' schooling "bought" with a 100 U.S. dollar investment, namely 0.77. For boys, we assume a net enrollment gain of 3 percentage points. The implied cost of enrolling an additional boy in upper primary school was therefore about $\$2.51/0.03 = \$84$. The years of boys' schooling "bought" with a 100 U.S. dollar investment, is then about 1.20. Our cost-effectiveness estimates fall within the range of estimates reported for other education interventions in Dhaliwal, Duflo, Glennerster, and Tulloch (2011) and Levy, Sloan, Linden, and Kazianga (2013).

2.2 Levy, Sloan, Linden, and Kazianga (2013)

The government intervention constructed one new mixed primary school with a canteen and free learning materials in each of the 132 treatment villages in Burkina Faso. Girls benefited in addition from take-home food rations, mobilization campaigns (information sessions, gender sensitization training to officials and teachers, etc.), training in literacy camps, and some girls' toilets and child care centers. Using a similar approach as above, the authors estimate that the annual cost of increasing village-level enrollment by one child was about \$70 in 2007 prices. This is the authors' estimate for constructing new program schools (rather than upgrading government schools with BRIGHT amenities) under the low-cost estimate for government schools.

2.3 Barrera-Osorio, Blakeslee, Hoover, Linden, and Raju (2011)

In a set of randomly chosen villages of Pakistan, for-profit entrepreneurs who were interested in establishing and running private primary schools were offered a per-child enrollment subsidy from the provincial government in 2009. In 100 villages, the subsidy was higher for girls (450 Rupees per month) than for boys (350 Rupees).

We impose a few simplifying assumptions to compute the cost-effectiveness of this program. First, the setup costs for infrastructure of program schools were entirely borne by the private provider. Second, students switching from public non-program to private program schools did not entail any change in running costs, which we assume to be equal for both school types. Furthermore, if all enrollment gains were concentrated in program schools, the only net expenditure borne by the government was the subsidy for previously not enrolled students. The annual costs of enrolling one more girl are hence just the annual subsidy, 12×450 Rupees, in 2009 prices. With an average nominal exchange of about 80 Pakistani Rupees per U.S. dollar during 2009, and accounting for U.S. inflation using the GDP deflator, this is equivalent to about 66 U.S. dollars.

2.4 Kim, Alderman, and Orazem (1999)

In 10 randomly selected slum neighborhoods in the city of Quetta, Pakistan, founders of new private schools were offered a subsidy for each enrolled girl, but not for boys, from early 1995. The nominal subsidy was 200 Rupees per girl and year, plus 150 Rupees per girl and month in the first year, 135 Rupees in the second, and 100 Rupees in the third. Using GDP deflators for 1995-97, the resulting annual average of 1,606 Rupees in 1995 prices is equivalent to the long-term costs of enrolling one additional girl, provided that the enrollment gain cannot be maintained without subsidies in the future and, as before, the net increase in the total expenditure on public education stems only from enrolling girls previously out of school. Given that prices increased by 132.6 percent between 1996 and 2005, an average nominal exchange of about 60 Pakistani Rupees per U.S. dollar during 2005, and accounting for U.S. inflation, the annual costs for getting one more girl into school is about 66 U.S. dollars in 2007 prices.

2.5 Banerjee, Jacob, Kremer, Lanjouw, and Lanjouw (2004)

In 21 randomly selected non-formal education camps in rural Rajasthan, India, the intervention sought to hire a second teacher, if possible female, in 1997. The authors calculate that the costs of increasing actual attendance by one month and for one child were 173 Rupees. The intervention affected only girls' attendance significantly. Prices increased by 45.5 percent until 2005, and with 45 Indian Rupees per U.S. dollar in 2005, and accounting for U.S. inflation, the amount needed to increase female attendance during one year was thus approximately 71 2007 U.S. dollars.

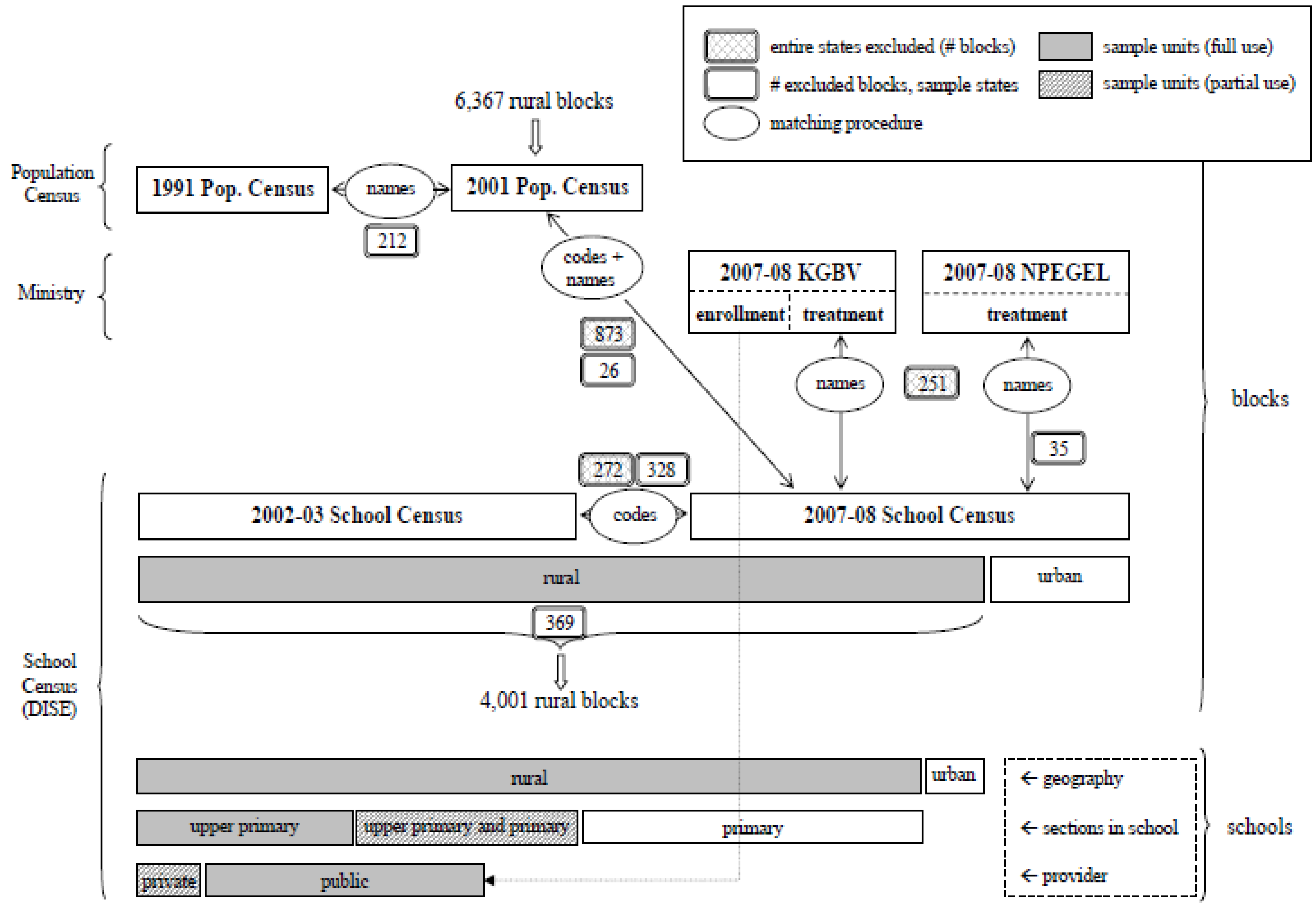

FIGURE A1. CREATION OF THE DATASET.