Impact of early childhood school intervention on enrolment and learning outcomes:

Evidence from a government-run pre-primary program in the Indian state of West Bengal

Nilanjan Bhattacharya♦, Debayan Pakrashi♥, Sarani Saha♠ and Soham Sahooℵ

[Preliminary draft, please do not cite]

This version: December 2022

Abstract

We evaluate the impact of the introduction of a pre-primary schooling program implemented in the government schools in the Indian state of West Bengal in 2013 on children's early enrolment in schools and subsequent test scores. Using double difference, triple difference and synthetic control methodology, we find that the program resulted in a significant increase in enrolment in the pre-primary sections of the government schools. However, the increase in enrolment did not translate into improved performance of the students. Analysing the test scores, we find that after the introduction of the program both math and reading scores for children in the state deteriorated in comparison to the neighbouring states. We attribute this result to the deteriorating physical and learning infrastructure in the state government schools with pre-primary sections, captured via decline in the average number of classrooms, average number of pre-primary teachers and decline in average teacher-student ratio.

Keywords: pre-primary education; learning outcomes, enrolment

JEL Classification: I21, I25, I28

♦ Department of Economic Sciences, Indian Institute of Technology Kanpur. Email: nilanjanb1@gmail.com
♥ Department of Economic Sciences, Indian Institute of Technology Kanpur and Economic Research Unit, Indian Statistical Institute Kolkata. Email: pakrashide@gmail.com
♠ Department of Economic Sciences, Indian Institute of Technology Kanpur, India. Email: sarani@iitk.ac.in
ℵ Centre for Public Policy, Indian Institute of Management Bangalore, India. Email: soham.sahoo@iimb.ac.in

1. Introduction

Good quality early childhood education (ECE) is pivotal for improving equitable education and lifelong learning opportunities for all. ECE can support better early learning outcomes (Britto et al., 2017; Berlinski et al., 2009) improved health outcomes (Elango et al., 2015), and even better social and economic development reaching into adulthood (Gertler et al., 2014). Within an education system, the benefits of ECE can translate into more equitable educational outcomes for marginalized groups (Berlinski et al., 2008). Given the benefits, it is not surprising that there has been rapid growth in ECE programs across the world in the last two decades (Behrman & Urzúa, 2013; Cascio, 2015; Nores & Barnett, 2010; Sayre et al., 2015; Wotipka et al., 2016). The provision of quality early childhood development, care, and pre-primary education to all children by 2030 as envisioned in target 4.2 of the Sustainable Development Goal 4 (SDG 4) also echoes its importance.

Despite the high return to early childhood investment in education, about fifty percent of children in the pre-primary age group (children aged between 3 and 6 years), which adds up to at least 175 million are deprived of pre-primary education globally (UNICEF 2019). In low-income countries, only one in every five children has access to pre-primary education (UNICEF 2019).[1] This failure limits children's futures, by denying them opportunities to reach their full potential, and it deepens inequities in later learning. The analysis of an ECE intervention in a low and middle-income context is therefore essential as a high proportion of the world's children reside in these countries.

In this paper, we contribute to the growing evidence base by evaluating the causal impact of a government-run free pre-primary program implemented in 2013 on children's test scores and enrolment in the Indian state of West Bengal. The issue of the impact of a pre-primary program becomes far more relevant in a setting like India where quality ECE programs are not available to millions of young children, particularly children from socio-economically disadvantaged backgrounds. To explore the pathway, we also analyze if the physical and learning infrastructure of schools has any role to play that ultimately affects the learning outcomes of the children. By doing so, we provide suggestive evidence of the short-run impact of ECE participation in the context of low and middle-income countries. Our study is also in view of the recently announced 'New Education Policy 2020' (NEP 2020) of India. Currently, children in the age group of 3-6 are not covered in the 10+2 education structure in India as class 1 begins

[1] https://www.unicef.org/media/57926/file/A-world-ready-to-learn-advocacy-brief-2019.pdf

at age 6. In the new proposed 5+3+3+4 structure of the NEP 2020, a strong base of Early Childhood Education (ECE) from age 3 is included, which is aimed at promoting better overall learning, development, and well-being. By focussing on a government-run free pre-primary program, we analyze the effectiveness of universal provisioning of quality early childhood development as envisioned in the NEP 2020.

In particular, we analyze the causal impact of a government-run, pre-primary program on enrolment and learning outcomes by exploiting the exogenous variation in exposure to pre-primary schooling brought about by the implementation of the pre-primary program in 2013. We employ both double-difference and triple-difference methodologies to estimate the effects on test scores and enrolment. We have also used the synthetic control methodology to rule out any bias in results that can occur due to the ad-hoc nature of the selected states in triple-difference regressions.

Our main findings suggest that the program has been quite successful in increasing enrolment but has resulted in a deterioration in learning outcomes captured via math and reading scores. Results from the difference-in-difference show that the program increased the number of government schools with a dedicated pre-primary section by 57 percentage points between pre and post-2013 as compared to the private schools in West Bengal. A similar result for the enrolment of pre-primary students has been found. The change in pre-primary enrolment in the government schools in West Bengal showed a massive 137 percentage point increase after the program was introduced in 2013 when compared to their private counterparts. When we compare West Bengal to its neighbouring states of Bihar, Jharkhand, and Orissa, using the triple difference method, the picture remains the same. Both availability and enrolment in the pre-primary section of government schools in West Bengal reported a huge increase in numbers that is statistically highly significant.

The success story of enrolment, however, did not translate into better learning outcomes. A comparison of pre-primary children between the treated and control cohort using the double difference method shows that treated cohort children from government schools in West Bengal do not achieve any better learning skills when compared to the control cohort. The results, however, change when we bring the neighbouring states into the picture. A comparison between West Bengal and its neighbouring states shows that government school children from West Bengal perform poorly both in math and reading tests than children from the same cohort in the neighbouring states.

To rule out any bias due to the choice of neighbouring states, we also made use of the synthetic control methodology. We find that both availability and enrolment in the pre-primary section in government schools experienced a massive increase after the program was implemented in 2013. However, math and reading scores deteriorated in government schools in West Bengal when compared to its synthetic control counterpart. This result is in accordance with the previous findings. We attribute this failure to the worsening physical and learning infrastructure of the state government-run schools in West Bengal. Comparing the number of classrooms, number of pre-primary teachers, and teacher-student ratio between the government schools in West Bengal to its neighbouring states, we find that all these indicators deteriorated in government schools of West Bengal. So, the school infrastructure did not complement the higher enrolment, leading to overcrowding in government schools.

Two core pieces of evidence emerge from the existing literature justifying the focus on ECE programs. Firstly, research shows that rates of long-term returns are high (between 5% to 14%) if the investment is made in ECE programs (Heckman et al., 2010; Garcia et al., 2017). Secondly, socio-economically disadvantaged children benefit the most from early childhood intervention (Heckman, 2013; Elango et al., 2015). These findings have given rise to the concept of dynamic complementarity in skill formation. Children who benefit from early human capital investments may benefit more from later investments (Cunha and Heckman, 2007). However, these success stories may have an ambiguous impact in the context of low and middle-income countries. Our paper contributes to the growing literature on early childhood education in the context of low and middle-income countries in several important ways. First, one of the limitations of the previous studies is that the evidence of strong results from early childhood interventions typically comes from high-dosage, holistic ECE programs, which differ substantially from the ECE programs that low- and middle-income countries have to offer (Behrman & Urzúa, 2013). In our paper, we try to find the shreds of evidence by focussing on a government-run free pre-primary program in a lower-middle income developing country setting. Second, some of the evidence is based on a handful of randomized-control trials that had small sample sizes, rather than regionally or nationally representative data that would permit generalizations (Heckman, 2011). We get rid of this limitation by using the government-administered DISE (District Information System for Education) and ASER (Annual Status for Education Report) data which are nationally representative databases with large sample sizes. Third, the populations in existing studies are often targeted and socioeconomically disadvantaged, which raises the question of whether similar benefits could be achieved within a general population (Baker, 2011). In our paper, the program in question has a free and universal rollout and is aimed at the general population. The significant contribution of this paper, therefore, lies in the evaluation of the government's pre-primary program, which is to the best of our knowledge the first study to causally evaluate a specific early childhood government intervention in the context of India and suggest appropriate policy reforms.

The rest of the paper is organized as follows. Section 2 describes the program in detail followed by data and some descriptive statistics in Section 3. The empirical strategy used in the paper is discussed in detail in Section 4. Next, we discuss the results of the paper estimated using the above-mentioned specification in Section 5. In section 6, we perform some robustness checks to reaffirm our results. We further analyse the possible reasons for the obtained results in Section 7 and finally conclude the paper with Section 8.

2. The Description of the Program

The education system of India is one of the largest in the world with more than 1.5 million schools, 8.5 million teachers, and 250 million students (UNICEF 2018).[2] However, there is a downside to the story when it comes to learning proficiency as about half of the children enrolled in primary schools, constituting nearly 50 million – could not achieve grade-appropriate learning levels (National Achievement Survey, NCERT 2017)[3]. Furthermore, children's school readiness has been far below the expected levels. For example, the percentage of all children in Std III who could read at a level appropriate for Std II was only 27.2% in 2018 (ASER 2018).[4] This learning crisis is largely common in the initial years of schooling. Most children enter primary school without any prior preparation. In addition, there is widespread inequity in the education system of India due to socio-economic factors that are of particular interest. In rural India, for example, by the age of seven, an achievement gap already exists between children from lower income groups, who are first-generation learners, and their comparatively wealthier counterparts with educated parents (Alcott and Rose, 2017). A quality early childhood education program can close down this learning gap.

[2] https://www.unicef.org/india/media/2596/file/Catalysing-transformational-change-in-school-education.pdf
[3] https://ncert.nic.in/pdf/NAS/WithReleaseDate_NPPTL.pdf
[4] http://img.asercentre.org/docs/ASER%202018/Release%20Material/aserreport2018.pdf

Recognizing its importance, section 11 of the 'The Right of Children to Free and Compulsory Education Act, 2009[5]' of India states that "*with a view to prepare children above the age of three years for elementary education and to provide early childhood care and education for all children until they complete the age of six years, the appropriate government may make necessary arrangement for providing free pre-school education for such children*". The national legal framework provided guarantee through this act and the government of West Bengal (WB) introduced a free one-year pre-primary education in government schools in the academic session 2013[6]. According to the new rules, a student whose age is between 5 and 6 years on the first day of the academic session (i.e., on 1st of January 2013) wiould be eligible to take admission in the pre-primary section. Apart from the pre-primary section, the age criteria for admission to different classes were also revised. Before 2013, five-year-old children could get admission in grade 1.[7] However, starting in 2013, only six-year-old students could enrol in grade 1 of the government schools.[8]

The program guidelines also mentioned that separate seating arrangements should be made available to the pre-primary students as far as possible. If due to the unavailability of space, schools are unable to accommodate them in a separate classroom, they could sit with grade 1 students. Their teaching and learning process should be carried out by the existing teachers in government schools. The pre-primary students are also entitled to receive benefits under the mid-day meal scheme. After having the mid-day meal, pre-primary students are allowed to leave the school.

3. Data & Descriptive Statistics

We use data from the Annual Status of Education Report (ASER), a yearly survey conducted to assess the status of education among children in almost all the rural districts of India from 2009 to 2018. The survey covers a random sample of about 20–30 households from each of the 20 villages selected from each of about 550 rural districts of India. From each household surveyed, all children in the age group 3 to 16 are surveyed, and the learning outcomes of children in the age group 5 to 16 are assessed.

[5] https://legislative.gov.in/sites/default/files/The%20Right%20of%20Children%20to%20Free%20and%20Compulsory%20Education%20Act,%202009.pdf
[6] https://wbxpress.com/files/2012/11/Admission_Age.pdf
[7] http://wbbse.org/Files/s_210_15022012.pdf
[8] https://wbxpress.com/files/2012/07/Age_Admission.pdf

The survey gathers detailed information on basic arithmetic and reading proficiency levels using well-tested rigorous tools. These tools are administered to all children across the districts and states and have been used extensively by other studies (see Chakraborty & Jayaraman, 2019; Lahoti & Sahoo, 2020). In the ASER data, the assessment of reading skills has four ordinal levels—recognition of letters, reading of words, reading a short paragraph (a grade 1 level text), and reading a short story (a grade 2 level text). On the other hand, the arithmetic skill assessment comprises four levels—recognition of single-digit numbers, recognition of double-digit numbers, subtraction of two-digit numbers with a carry-over, and division of a three-digit number by one digit.

In our analysis, we have considered raw scores as a measure of learning outcomes. We did not use class detention or promotion rates as a measure of learning abilities because the standards for promotion depend on the standard of tests conducted at the school level. It is possible that schools with well-performing students conduct learning tests that are of a higher standard in comparison to other schools with supposedly low achievers. As the tests conducted by the schools vary widely across schools, they cannot be treated as standardised; therefore, we used assessments conducted by ASER, which are uniform across the nation. This has been pointed out in the literature pertaining to educational research (Anaya, 1999). Apart from the variables on learning, the ASER survey also collects child, household, and village-level information that can be used as independent variables in the regressions and are possible confounders of learning outcomes. Household economic characteristics were controlled through a number of indicators: whether the house was cemented or not; whether it had electricity; possession of a toilet; possession of a television; and total members within the household. Child-level characteristics include the child's age. Village-level factors controlled for include whether the village has a private school, a private health clinic, a bank or a cemented road. Controlling for these characteristics enabled us to get an unbiased estimate of the impact of the pre-primary program.

However, as ASER does not include any information on whether the child in question is enrolled in the pre-primary section of a school or not, it is not possible to identify the pre-primary education status of the children. To analyse the effect of the program on enrolment at the pre-primary level, we have made use of the annual District Information System for Education (DISE) database. The DISE dataset gathers detailed information on different school-level characteristics ranging from school infrastructure, facilities, enrolment, and teachers for all the districts. In our analysis, we have used DISE data from the years 2009 to 2017.

To begin, we use the enrolment and infrastructure data from DISE to see if the implementation of the program in 2013 increased the number of government schools with a pre-primary section and the number of students enrolled in the pre-primary section in West Bengal. See Figures 1 & 2.

[Insert Figure 1]

[Insert Figure 2]

Interestingly, we notice an immediate jump in both the number of government schools with a pre-primary section and the number of students enrolled in pre-primary for the year 2013, when this pre-primary program was implemented. While the number of government schools with pre-primary drastically increased to almost 60,000, the total number of students enrolled in the pre-primary sections of these government schools recorded a figure of almost 800,000 in 2013. We did not observe any such change in the private schools. To verify that similar changes were not taking place in other states during that same period, we compare the number of government and private schools with pre-primary section (and students enrolled in these schools) in West Bengal to the same in the neighbouring states of Bihar, Jharkhand, and Orissa. We consider this group of states as a comparison group to West Bengal since there are substantial similarities in terms of social, cultural, and economic conditions between the states. Further, it must be noted that the average pre-primary availability, reading, and mathematics skills in West Bengal (WB) before the implementation of the program (2009-2012) were similar to the same in the other states (see Figure 3).

[Insert Figure 3]

[Insert Figure 4 & Figure 5]

Figures 4 & 5 suggest that the massive expansion of pre-primary education in government schools post-2013 took place only in the state of West Bengal as compared to the neighbouring states considered in the analysis. Combined together, Figures 1-4 are the stepping stones that further motivate us to dig deeper into estimating the causal impact of the pre-primary program.

4. Empirical Methodology

Our study exploits the exogenous variation in exposure to pre-primary schooling and makes the identification of the causal impact plausible[9]. In particular, we apply a difference-in-difference (DD) approach to estimate the causal effect of the program on pre-primary enrolment and learning outcomes. We use DISE data to assess the impact of the implementation of the program on enrolment at the school level and rely on ASER data while focusing on the learning outcomes at the individual level.

A major challenge in identifying the causal effect of pre-primary school attendance on learning outcomes is the non-random selection of children into early education. Positive selection, whereby parents whose children attend pre-primary school possess characteristics that promote better school performance, would result in a spurious positive correlation between pre-primary and later academic outcomes. Since children are not randomly selected for pre-primary education, selection based on parental heterogeneity is likely to be unavoidable. However, as the program in discussion over here has a universal roll-out at the state level and the age at entry does not change, the question of motivated parents sending their children for pre-primary education does not arise. The chance of positive selection bias is, therefore, ruled out. We hypothesize that this one year of extra schooling in the form of pre-primary education for government-run schools can have a positive impact on these children's academic achievement, measured using the standardized literacy and numeracy scores.

4.1. Difference-in Difference

Considering that DISE is the source of the enrolment data, the government schools by virtue of being exposed to the program, are considered as the treated group whereas the private schools which are not exposed to the program are included in the control group. For learning outcome, test scores for 5-6 years old children are the dependent variables. The first difference compares the desired dependent variables across these two types of schools in West Bengal. However, this difference is likely to be confounded due to several other changes occurring in West Bengal over time and we cannot attribute this first difference as the impact of the pre-primary program. To eliminate the impact of time trend on the results, the difference across private and government schools of West Bengal after the implementation of the program (i.e.,

[9] Prior to the implementation of the pre-primary schooling program, the school entry age for government school students was five years for grade I. After the introduction of the program, the entry age for children in schools remained the same but now these five-year-old children were eligible to be enrolled in the pre-primary section of the schools only and would enjoy a year of pre-primary education before they are promoted to grade I at the age of six.

year 2013) is compared with the same types of schools before the program was implemented. The difference-in-difference (DD) equation to be estimated therefore is as follows:

$$(1)\quad y_{i\,dt} = \beta_0 + \beta_1\, govtschool_{id} + \beta_2\, post_t + \beta_3\, (govtschool_{id} X\, post_t) + \vartheta X_i + \gamma_t + \partial_d + (\partial_d\, X\, t) + \epsilon_{i\,dt}$$

The main specifications are the same for both sets of regressions for enrolment and learning outcomes. The only difference is that the dependent variables for enrolment are measured at the school level whereas the test scores for learning outcome are child level variables. Here, $Y_{i\,dt}$ is the outcome variable of interest (availability of pre-primary section in the school and number of students enrolled in pre-primary sections captured via logarithm of total number of pre-primary students) for the i-th school from district d measured at time period t. For learning outcomes, $Y_{i\,dt}$ corresponds to the test score (math and reading score) of the i-th child from district d at time t. The first outcome variable for enrolment—availability of pre-primary section— is a dummy that takes the value 0 if the i-th school does not have a pre-primary section and the value 1 if it has a pre-primary section while the second variable just focuses on the number of students enrolled in the primary sections of the i-th school.

For learning outcomes, $Y_{i\,dt}$ takes integer values from 0 to 4 where 0 means no learning skills and 4 implies the highest level of learning. ASER's tests comprise of 5 levels, which represent cumulative skill mastery for both the literacy as well as the numeracy assessments. For both tests separately, the child is marked at the highest level that he/she can do comfortably. The first measurement issue arises from the fact that we have assigned integer values (0,1,2,3 or 4) to these levels. In so doing, we are treating test scores as interval scales, when in fact, they have an ordinal scale. This is problematic for two reasons. First, comparisons across groups are sensitive to the choice of scale. Second, interpretation of the treatment effects are tricky in this context: with an ordinal scale, it is hard to know which levels of learning achievement are driving the percentage improvement over the baseline test scores. To address this issue, we have considered two sets of dependent variables throughout the paper when estimating the impact on learning outcome. The first set of outcome variables report the raw scores from both the math as well as the reading test while the second set report the age-wise standardized test score as the dependent variable.

The dummy ‘ $govtschool_{id}$’ on the other hand takes the value 1 if the i-th school is a public school run by the government and a value 0 if it is a private school. Similarly, for regressions on learning outcomes, ' $govtschool_{id}$' takes the value 1 if the i-th child from district d goes to a government school and 0 otherwise. The '$post_t$' dummy compares the outcomes for the post program years (2013 or after) when it assumes the value 1 to the same before 2013 (for '$post_t$'=0). We are particularly interested in the sign and magnitude of the coefficient β_3 , which is the causal intent-to-treat (ITT) estimate of the impact of the pre-primary program on the outcome variables. All the child-, household-, school- and village level characteristics have been clubbed under X_i. For school level regression on enrolment, X_i includes the total number of classrooms, the total number of teachers, and availability of electricity in the school.

For child level regression on learning outcomes, X_i controls for the household variables for the i-th child that include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics that include the child's age, whether the mother went to school and village-level factors such as whether the village has a government primary school, private school, village post office, bank, pucca road in the village, are also controlled under X_i. Finally, we include district (∂_d), year (γ_t) fixed effects and district specific linear trend (∂_dx t) for both sets of regressions.

To validate the DD estimator, it is, however, necessary to establish that the trend in the outcome variables were the same for the treated and untreated schools in the absence of treatment (following Angrist and Pischke, 2009). This is the key identification assumption of DD and is known as the parallel-trend assumption. Although it is not possible to test this assumption – since we cannot observe the treated group in the absence of treatment– we can get some indication about this validity by checking the pre-treatment trends. If the enrolment trends of the government and private schools evolved in similar ways before the program, they would be expected to continue to do so in the post periods in the absence of the pre-primary program. To test the parallel trend assumption, we therefore take data from the years 2009 to 2012 and run the regression equation (1) with 2010 as the reference year for the 'time' dummy. The 'time' dummy thus takes the value 1 for years (2010-2012) and 0 otherwise. An insignificant β_3 here would imply that the parallel trend assumption holds.

In the previous regression, the double difference consisted of two dimensions— the type of school and time. However, there is also a possibility that the government schools in West Bengal are significantly different from the government schools in the neighbouring states. To evaluate this change, we replace the ' $govtschool_{id}$' variable from equation 1 by a dummy variable for West Bengal and run the following regression.

$$(2)\ y_{idst} = \beta_0 + \beta_1 . post_t + \beta_2 . (post_t\ X\ wb_s) + \vartheta X_i + \partial_{ds} + \gamma_t + (\partial_{ds} X\ t) + \epsilon_{idst}$$

In this equation, the double difference consists of both time and state dimensions. For enrolment, y_{idst} denotes the availability of pre-primary, log (1+number of pre-primary students) in the *i*-th school from district *d* of state *s* at time *t*. For regression on learning outcomes, y_{idst} stands for the test score of the *i*-th children from district *d* of state *s* at time *t*. The remaining variables are same as in equation 1 except 'wb_s' that takes the value 1 if the observations are from West Bengal and 0 for neighbouring states. Again, we need to validate the DD estimator by testing the parallel trend assumption. For this purpose, we run the regression equation (2) by restricting our sample from 2009 to 2012 and replacing the '$post_t$' dummy by '$time_t$' dummy.

4.2 Triple Difference

Even though the DD estimates give us an estimate of the impact of the program on enrolment, one might still argue that there can be certain other unobserved factors like the general equilibrium effects or broader trends affecting the DD estimate. Thus, a Triple difference (TD) or Difference-in-difference-in-differences approach (DDD) needs to be adopted to nullify any other broader trends that might bias our DD result. Following Muralidharan & Prakash (2017), we make use of a Triple Difference (TD) regression by comparing the DD estimates from West Bengal to a bunch of neighbouring states including Bihar, Orissa, and Jharkhand.

Once we have successfully chosen the control states, we are able to move on to estimate the impact of the program on pre-primary enrolment and learning outcomes by comparing West Bengal to the above-mentioned control states. We estimate the following Triple-difference (TD) model:

$$\begin{aligned}(3)\ \ y_{idst} = {} & \beta_0 + \beta_1 .\ govtschool_{ids} + \beta_2 . post_t + \beta_3 . (wb_s\ X\ govtschool_{ids} X post_t) \\ & + \beta_4 . (govtschool_{ids} X\ post_t) + \beta_5 . (\ post_t\ X\ wb_s) + \beta_6 . (govtschool_{ids}\ X\ wb_s) + \vartheta X_i \\ & + \gamma_t + \partial_{ds} + (\partial_{ds}\ X\ t) + \epsilon_{i\,dst}\end{aligned}$$

Here, the 'wb_s' dummy takes the value 1 if the school in question is in the state of West Bengal and 0 otherwise. The rest of the variables used in equation (3) has already been discussed while discussing the DD model. In this model, we are particularly interested in the sign and magnitude of the β_3 , which captures the causal ITT estimate of the impact of the pre-primary program on the outcome variables. In the TD model, we again check the parallel trend assumption as has been done earlier for the DD approach.

4.3 Synthetic Control

One limitation of our TD regression analysis is that the control states, which were chosen because of their geographical proximity, cultural, and socioeconomic similarities, may not be suitable and it could thus be argued that this choice was arbitrary and ad hoc. To ensure that our estimates are unbiased and avoid arbitrary choice of control states, we have used synthetic control method (SCM) that has been used in extant literature to measure impact (Abadie & Gardeazabal, 2003; Abadie et al, 2010; Peri & Yesenov, 2019). These papers argue that SCM is methodologically superior to the classic TD because it takes a linear combination of states that is found to form a better control group for WB than a single state. More specifically, we created a synthetic optimal control group[10], which minimised the pre- program difference with WB for a given set of relevant characteristics that determine educational outcomes. The synthetic state here reduces the ad hoc nature of the choice of the control states, which were otherwise chosen without applying the SCM methodology. For this, we made use of the ASER and DISE database from 2009 to 2018. We calculated the state-level annual estimates of learning outcomes among Government school children in the age group of 5-6 years[11] by using the ASER data.

To obtain the synthetic control estimates for the pre-primary enrolment, we have used state-level annual estimates of the proportion of schools with pre-primary sections and the total number of students in pre-primary sections by using DISE data. The predictors of the pre-primary enrolment for the synthetic control are the state-wise average number of classrooms in government schools, proportion of government schools with electricity, and playground. For learning outcomes, predictor variables for test scores are the proportion of households that are fully cemented, average household size, proportion of households having a TV, proportion of households with electricity, proportion of households with toilet, proportion of mothers who

[10] The control states for the synthetic control method are Himachal Pradesh, Punjab, Uttarakhand, Haryana, Rajasthan, Uttar Pradesh, Bihar, Sikkim, Nagaland, Manipur, Mizoram, Tripura, Meghalaya, Assam, Jharkhand, Orissa, Chhattisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Karnataka, Kerala and Tamilnadu.

[11] A child is admitted in pre-primary when he is aged between 5 and 6 years. Therefore, children in the cohort of 5 and 6 years of age are exposed to the pre-primary program.

went to school at some point of time, proportion of villages with pucca road, proportion of villages with a bank, proportion of villages with post office, proportion of villages with private school and proportion of villages with government primary school. The outcome variables for the period 2009 to 2012 are also used as predictors.[12]

5. Results

In this section of the paper, we present the results related to the impact of the pre-primary program on first enrolment in Section 5.1 and then learning outcomes in Section 5.2.

5.1 The impact on enrolment

We begin this section, with the results from the DD regression estimated using Equation 1 and presented in Table 1. The coefficient β_3 associated with the interaction term between 'Government school' and 'Post' is positive and highly significant for both the dependent variables considered, with a 58 percentage points increase in the probability of a government school in WB having a pre-primary section and 139 percentage points increase in enrolment in the pre-primary section of the government schools, between pre- and post-2013 as compared to the private schools.

[Insert Table 1 here]

The parallel trend result using the DISE data from 2009 to 2012 is presented in appendix Table 1. The significant coefficient of the interaction term for the regression on availability of pre-primary implies that the null hypothesis of parallel trend can be rejected. This further means that the difference between the treatment and control group (in our case government schools and private schools respectively) in terms of availability of pre-primary schools before the program would not remain same had the treated group (government schools) not been exposed to the program. So, we move on to the other DD specification.

Next, we investigated if the government schools in West Bengal are significantly different from the government schools in the neighbouring states in terms of availability of pre-primary sections and enroment in it. We estimated equation 2 and present the results in Table 2.

[Insert Table 2 here]

[12] The weights for the individual states and predictor balance between the treated and the synthetic group are reported in Appendix Table 1-4 for both enrolment and learning outcomes.

The coefficient β_2 associated with the interaction term between 'wb_s' and '$post_t$' dummy is positive and highly significant for both the outcome variables. The result shows that the probability of having a pre-primary section in government schools in West Bengal increased by more than 90 percent as compared to the government schools in the neighbouring states. On the other hand, enrolment in pre-primary sections increased two-fold times in the government schools in West Bengal in comparison to the same type of schools in the neighbouring states.
Again, we tested for the parallel trend assumption and the result in appendix Table 2 shows that the difference between the government schools in West Bengal and its neighbouring states before the program would remain the same had the pre-primary program not been implemented.

Finally, we employed the triple difference method to compare West Bengal with a bunch of neighbouring states. We investigate if the differences between government and private schools in West Bengal in terms of the availability of pre-primary section and number of pre-primary students are significantly different from the same differences across these outcome variables between government and private schools in the neighbouring states.

[Insert Table 3 here]

We estimate equation 3 and the findings from Table 3 confirm that there is a positive significant impact of the pre-primary program on both the availability of pre-primary sections in government schools and the number of students enrolled in the pre-primary section. The difference between government and private schools in West Bengal in terms of the number of schools with pre-primary section resulted in an increase of 72 percentage points when compared to the same difference between government and private schools in the neighbouring states. Similarly, the difference between government and private schools in West Bengal in terms of the number of pre-primary students enrolled reported a 173 percentage points increase as compared to the same difference between government and private schools in the neighbouring states. Test for the parallel trend assumption for the TD specifications are reported in appendix Table 3 and the result shows that the parallel trend assumption holds. This means that the trends for the schools exposed to the pre-primary program were similar to that of the schools without the program to begin with. So, the results from both the DD and TD regressions suggest that pre-primary program was successful in increasing enrolment and availability of pre-primary education in the government-run schools. This establishes the positive impact of the program on enrolment.

5.2 The impact on learning outcomes

The series of regressions for learning outcomes starts with the most simplified one. First, we consider West Bengal and compare the government school children of pre-primary age group to the children from private schools of the same age-group. In doing so, we estimate the equation 1 for learning outcomes. The results presented in Table 4 suggest that the coefficient associated with the interaction term between '$govtschool_{id}$' and '$post_t$' is insignificant for both the test scores. This implies that reading and math scores of government school pre-primary students in West Bengal did not differ significantly from the private school children. We also tested for the validity of parallel trend assumption and the results from appendix Table 4 violate the assumption.

[Insert Table 4 here]

Next, we bring in the neighbouring states of West Bengal and investigate if the test scores of government school pre-primary children in the state are significantly different from the test scores of government school pre-primary children in its neighbouring states. The coefficients associated with the interaction term between '$post_t$' and 'wb_s' are found to be negative and statistically insignificant for both raw test scores in Table 5. So, government school students of WB in the pre-primary age group perform worse both in math and reading tests than children from the same category in neighbouring states.

[Insert Table 5 here]

The estimates from Table 5 show that the interaction term between '$post_t$' and 'wb_s' is negative and significant for both the test scores. It shows that the government school children in West Bengal perform poorly than the same group of children in the neighbouring states but the performance in math is worse than the performance in reading score. We also tested for the parallel trend assumption and found it to be valid as reported in appendix Table 5.

Finally, we combined West Bengal and the control states together and investigate if the difference in learning outcomes between the pre-primary children from government and private schools in West Bengal is significantly different from the difference in learning outcomes between the pre-primary children of government and private schools in the neighbouring states. Essentially, we run the TD regression as in equation 3.

[Insert Table 6 here]

Table 6 shows that the difference in test scores between government and private school pre-primary children in West Bengal is not significantly different from the difference in test scores between government and private school pre-primary children in the neigbouring states. Again we tested for the parallel trend assumption and found that the assumption is violated for math score as reported in appendix Table 6.

So, comparing both the equations of the DD and TD regression, we can say that equation 2 is our preferred specification for learning outcomes as the parallel trend assumption holds in this case only.

The major findings from the regressions on learning outcomes indicate that both math score and reading score have deteriorated after the implementation of the program if we compare West Bengal to its neighbouring states. The decrease in math score is more than the decrease in reading score.

5.3 Synthetic Controls

Results from the synthetic control reaffirms our findings from the previous analyses using DD and TD regression specification.

[Insert Figure 6 here]

[Insert Figure 7 here]

Figures 6 (a) and 6 (b) present the trends in the percentage of government schools with pre-primary and number of pre-primary students in government schools respectively. We see that there is a significant increase in pre-primary enrolment (both availability of pre-primary section and number of pre-primary students) following the implementation of the program in West Bengal in comparison with the synthetic state.

We also examined the effect on learning outcomes by using the SCM. We used the same set of predictors to generate a similar figure for reading and mathematical ability outcomes as defined earlier. Figure 7 (a) shows that math score has deteriorated in WB following the implementation of the program. From figure 7 (b) we see that except the year 2014 reading scores for 5-6 years old government school children in West Bengal have deteriorated.

Combined with the high degree of balance on all predictors, figure 6 (a), 6 (b) and figure 7 (a), 7 (b) suggest that the synthetic West Bengal provides a sensible approximation to both pre-primary enrolment and learning outcomes that would have prevailed in West Bengal post 2013 in the absence of the pre-primary program. Our estimate of the effect of the pre-primary program on enrolment and learning outcomes is the difference between the outcome variables (both enrolment and learning scores) in WB and in its synthetic version after the pre-primary program in 2013. Immediately after the implementation, the two lines of pre-primary availability and pre-primary enrolment begin to diverge noticeably. However, different pictures can be noticed for learning outcomes. Post 2013, the two lines of math score diverge slightly with the dotted line corresponding to the synthetic state lying above the line for WB, indicating a deterioration of math scores. Similar picture is demonstrated in the case of reading score. The line corresponding to WB lies below the line for the synthetic state for all the post treatment years except 2014, implying a deterioration of the reading score. The difference between the two lines (West Bengal and its synthetic control counterpart) is denoted as 'gap' which essentially gives us the effect of the program.

One of the limitations of the SCM is that bias may be present in the synthetic control estimated marginal treatment effects (and thus the estimated average treatment effect estimates) because of discrepancies between the predictor variable values in each treated unit and its synthetic control donors. A bias-correction procedure, analogous to the approach in Abadie and Imbens (2011) to address inexact matching on predictor variables with matching methods, has been proposed (Abadie and L'Hour, 2021; Ben-Michael et al., 2021). This does not necessarily improve the pre-treatment fit of the outcome variable, but rather addresses discrepancies between a treated unit and its donor pool in the values of all specified linear combinations of predictor variables, including the covariates. We call this later one as the 'Bias-corrected' estimates and the original one as the 'Classic' estimates (Abadie et al., 2010). Next, we present the 'gaps' or the effect graphs with both the 'Classic' and 'Bias' corrected estimates.[13]

[Insert Figure 8 here]

[Insert Figure 9 here]

Panel A and panel B in Figure 8 suggest that the program had a huge positive effect on both pre-primary availability and number of pre-primary students and that this effect increased over time for the number of students but remained steady for the availability of pre-primary schools. The magnitude of the estimated impact of both the graphs in panel A and panel B is substantial.

[13] These bias-corrected graphs are created using STATA's 'allsynth' command (Wiltshire; 2021).

Our results suggest that post 2013, both number of students and percentage of government schools with pre-primary increased significantly. Interestingly, both the graphs in panel A and panel B in figure 9 demonstrate that the effect of the pre-primary program on math score and learning score is very weak as it is very close to the zero mark.

To assess the significance of our estimates, we conduct a series of placebo studies by iteratively applying the synthetic control method used to estimate the effect of the pre-primary program in WB to every other state in the donor pool. In each iteration we reassign in our data the pre-primary intervention to one of the 24 control states, shifting WB to the donor pool. That is, we proceed as if one of the states in the donor pool would have implemented pre-primary program in 2013, instead of WB. We then compute the estimated effect associated with each placebo run. This iterative procedure provides us with a distribution of estimated gaps for the states where no intervention took place. Following graphs demonstrate the results of the placebo run.

[Insert Figure 10]

Panel A and Panel B in figure 10 display the results for the placebo test on pre-primary enrolment. The grey lines represent the gap associated with each of the 24 runs of the test. That is, the grey lines show the difference in percentage of schools with pre-primary (Panel A) and number of students in pre-primary (Panel B) between each state in the donor pool and its respective synthetic version. The superimposed black line denotes the gap estimated for West Bengal. As the figures make it apparent, the estimated gap for WB during the post 2013 period is unusually large relative to the distribution of the gaps for the states in the donor pool.

[Insert Figure 11]

Panel A and Panel B in figure 11 represent the same results for the placebo test on learning outcome. It is evident from the figure that the estimated gap in both math and reading score for WB during the post 2013 period oscillate around the zero mark which is more or less similar relative to the distribution of the gaps for the states in the donor pool. So, the pre-primary program had a negative effect on learning outcomes.

6. Robustness check: Event study

Previous analyses of DD regression and synthetic control have established a positive significant impact of the pre-primary program on enrolment. However, we found negative impact on test scores. In this section we undertake an event study exercise for robustness check of the program's impact on both enrolment and learning outcomes. For enrolment, we run the regression as specified in equation 2 but with a slight modification. Instead of the dummy '$post_t$', we take different years and interact with the dummy for West Bengal. We document how avalibaility of pre-primary and enrolment evolved in government schools in West Bengal in comparison to the neighbouring states. The coefficient of the interaction term is plotted against different years in figure 12.

[Insert Figure 12]

From figure 12, we find a systematic pattern of change in the availability of pre-primary and student enrolment. Considering 2013 as the reference, the coefficient of the interaction term is positive and significant for both the outcome variables 2013 onwards whereas it is negative for years before 2013. It shows that both enrolment and availability of pre-primary increased significantly from 2013.

For learning outcomes, a different picture comes out from the event study graphs. We adopted the same strategy and interacted different years with the dummy for West Bengal for both the outcome variables. Both math and reading scores deteriorated in post program period except the years 2009 and 2010. The math and reading scores for 2009 and 2010 are not significantly different from the post program years. The decline in math score is more than the decline in reading score. Observations from figure 12 and figure 13 are in tune with our results from the previous analyses.

[Insert Figure 13]

7. Potential Mechanism for the Decline in Learning Outcome: School Infrastructure

Our findings related to a negative impact on test scores raise several questions. What are the possible reasons for this decline, given the context of West Bengal? We speculate that the detrimental impact on learning outcomes might have arisen as a result of structural factors like school facilities. For instance, if schools are inadequately staffed, increased enrolment might result in congestion externality and could negatively affect learning outcomes.

Using the DISE data, we compared some of the indicators of the condition of school infrastructure in government schools with pre-primary sections in West Bengal to the same category of schools in its neighbouring states. More specifically, we looked at both physical infrastructures, like the average number of classrooms, and learning infrastructure, like average number of teachers in pre-primary section and average teacher-student ratio. Figure 14 (Panel A– Panel C) presents the relevant graphs from 2009 to 2017.

[Insert Figure 14]

Correspondingly, we used equation 2 as that is our preferred specification, and re-run the set of regressions only for government schools with pre-primary. The outcome variables were logarithm of total number of classrooms, logarithm of number of teachers in pre-primary and teacher-student ratio. Results from Table 7 represent that all the three indicators declined in West Bengal as compared to the neighbouring states after the introduction of the program.

[Insert Table 7]

From figure 14, we observe for all the indicators that conditions in government schools with pre-primary in West Bengal have deteriorated in comparison with the government schools in the neighbouring state except for the year 2017. There can be certain state specific reasons behind the convergence of infrastructure indicators for the year 2017. The choice of the neighbouring states might be a factor also. So, we adopted the synthetic control methodology again to rule out any bias in our results. Results from the synthetic control methodology are presented in figure 15.

[Insert Figure 15]

It is evident from figure 15 (Panel A-Panel C) that both physical and learning infrastructure have deteriorated more in the government schools of West Bengal after the introduction of pre-primary. Panel B represents one of the most alarming pictures. The average number of teachers appointed for pre-primary section is not even 1 in government schools of West Bengal. This shows that although enrolment has increased but there is hardly any dedicated teacher for pre-primary students in government schools. This is the crude reality in the government schools of West Bengal. The pre-primary students sit with students from higher standards. They do not receive any extra attention from the teachers. As a result, they fail to develop any learning skills even though they enjoy one extra year of schooling in the form of pre-primary before being promoted to class 1.

8. Conclusion

We have added to the growing evidence on universal pre-primary programmes by estimating the impact on enrolment and education outcomes. Our estimates of the program indicate that it has a huge role in increasing pre-primary enrolment, hence the program is successful in sending children at their nascent age to schools. While this effect is prevalent across the state, we also observe that the program has a negative impact on the test scores of children.

Despite the positive influence on enrolment, we found a marked decline in learning ability (both math score and reading score). This deterioration can be attributed to the declining physical and learning infrastructure that has taken place in West Bengal over the years after the program was announced. The announcements made by the state have not been reflected in the improvement of school infrastructure. Extant literature shows that student achievement responds to the quality of the learning environment and that overcrowding in class has a negative effect on both teaching and learning (Corcoran et al., 1988; Angrist & Lavy, 1999; Case & Deaton, 1999). Keeping in mind the falling performance of the students especially in mathematics, the decreased teacher-student ratio and decreased average number of pre-primary teachers account for the deteriorated learning outcomes in government schools. Since, our findings indicate that this decline in learning skills is potentially the result of inadequate physical and learning infrastructure in West Bengal, the state may not have been able to cope with the increasing pre-primary enrolment arising from the programme.

If the sole objective of the pre-primary program was to make children familiar with the school environment and make it a habit for them to attend classes, our findings strongly support that and claim that the program has been instrumental in increasing pre-primary enrolment massively. However, if the benefits of pre-primary program in the form of better learning outcomes are to be reaped, we have many reasons to ponder upon. Improvement of school infrastructure is the first and foremost priority a benevolent government should have. From the point of view of the elected government, allocating funds for tangible assets that provide immediate returns assumes priority relative to, say, learning facilities, which have a long gestation period. Ensuring education quality involves costly screening activities and this might distort public investment. To add to this, as a populist measure a government might announce a free pre-primary program in an existing vulnerable, inefficient education system without investing in school infrastructure for larger political gain. Such announcements change the public perception of the government. As an instance, the government of West Bengal distributes school uniforms and cooked mid-day meals to pre-primary students. This is a welcome step and can be lauded for encouraging enrolment in pre-primary sections. However, such steps may be unlikely to enhance learning skills without complementary improvements in inputs pertaining to physical school infrastructure and numbers of teachers. Our paper shows that the gains in these inputs in West Bengal have been substantially lower in government schools in comparison with other states.

References

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American economic review*, *93*(1), 113-132.

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

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, *29*(1), 1-11

Abadie, A., & L’Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, *116*(536), 1817-1834.

Alcott, B., & Rose, P. (2017). Learning in India’s primary schools: How do disparities widen across the grades?. *International Journal of Educational Development*, *56*, 42-51.

Anaya, G. (1999). College impact on student learning: Comparing the use of self-reported gains, standardized test scores, and college grades. *Research in higher education*, *40*(5), 499-526.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly journal of economics*, *114*(2), 533-575.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baker, M. (2011). Innis Lecture: Universal early childhood interventions: what is the evidence base?. *Canadian Journal of Economics/Revue canadienne d'économique*, *44*(4), 1069-1105.

Behrman, J. R., & Urzúa, S. (2013). Economic perspectives on some important dimensions of early childhood development in developing countries. *Handbook of early childhood development research and its impact on global policy*, *23*.

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789-1803.

Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of public Economics*, *92*(5-6), 1416-1440.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of public Economics*, *93*(1-2), 219-234.

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., ... & Lancet Early Childhood Development Series Steering Committee. (2017). Nurturing care: promoting early childhood development. *The Lancet*, *389*(10064), 91-102.

Cascio, E. U. (2015). The promises and pitfalls of universal early education. *IZA World of Labor*.

Case, A., & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics*, *114*(3), 1047-1084.

Chakraborty, T., & Jayaraman, R. (2019). School feeding and learning achievement: evidence from India's midday meal program. *Journal of Development Economics*, *139*, 249-265.

Corcoran, T. B. (1988). *Working in urban schools*. Institute for Educational Leadership, 1001 Connecticut Avenue NW,# 310, Washington, DC 20036.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, *1*, 697-812.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). Early childhood education. In *Economics of Means-Tested Transfer Programs in the United States, Volume 2* (pp. 235-297). University of Chicago Press.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., ... & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, *344*(6187), 998-1001.

García, J. L., Heckman, J. J., Leaf, D. E., & Prados, M. J. (2017). *The life-cycle benefits of an influential early childhood program* (No. w22993). National Bureau of Economic Research.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics*, *94*(1-2), 114-128.

Heckman, J. J. (2011). The economics of inequality: The value of early childhood education. *American Educator*, *35*(1), 31.

Heckman, J. J. (2013). *Giving kids a fair chance*. Mit Press.

Lahoti, R., & Sahoo, S. (2020). Are educated leaders good for education? Evidence from India. *Journal of Economic Behavior & Organization*, *176*, 42-62.

Muralidharan, K., & Prakash, N. (2017). Cycling to school: Increasing secondary school enrollment for girls in India. *American Economic Journal: Applied Economics*, *9*(3), 321-50.

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world:(Under) Investing in the very young. *Economics of education review*, *29*(2), 271-282.

Peri, G., & Yasenov, V. (2019). The labor market effects of a refugee wave synthetic control method meets the mariel boatlift. *Journal of Human Resources*, *54*(2), 267-309.

Sayre, R. K., Devercelli, A. E., Neuman, M. J., & Wodon, Q. (2015). Investing in early childhood development: review of the World Bank's recent experience.

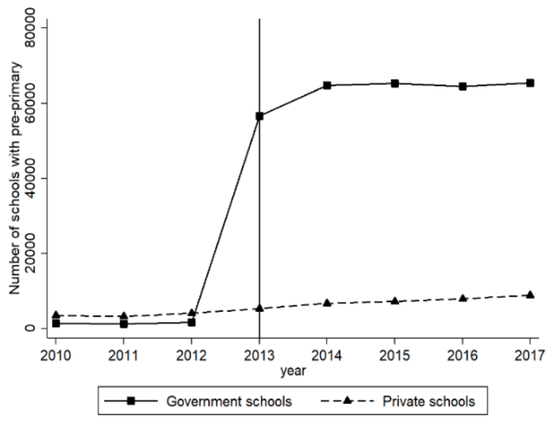
UNICEF. (2019). A World Ready to Learn: Prioritizing Quality Early Childhood Education. New York: UNICEF.

Wiltshire, J.C., 2022. allsynth: (Stacked) Synthetic Control Bias-Correction Utilities for Stata. Working paper. https://justinwiltshire.com/s/allsynth_Wiltshire.pdf.

Wotipka, C. M., Rabling, B. J., Sugawara, M., & Tongliemnak, P. (2017). The worldwide expansion of early childhood care and education, 1985–2010. *American Journal of Education*, *123*(2), 000-000.

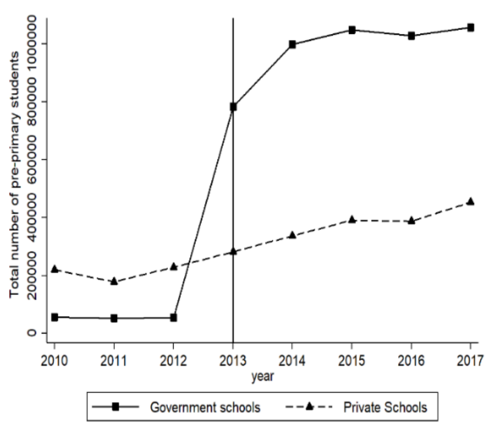
Figures

Figure 1: Number of schools with pre-primary in West Bengal



Notes: DISE data from 2010-2017 were used for schools in West Bengal only.

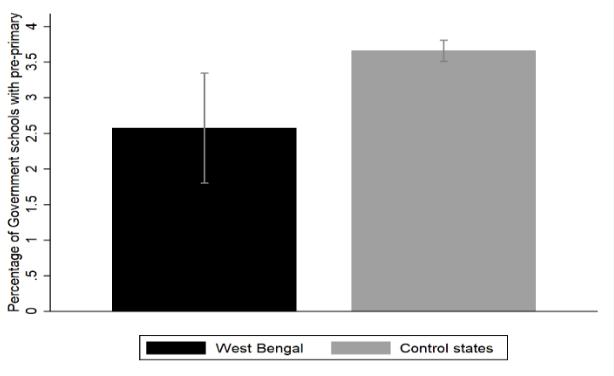
Figure 2: Student enrolment in pre-primary in West Bengal



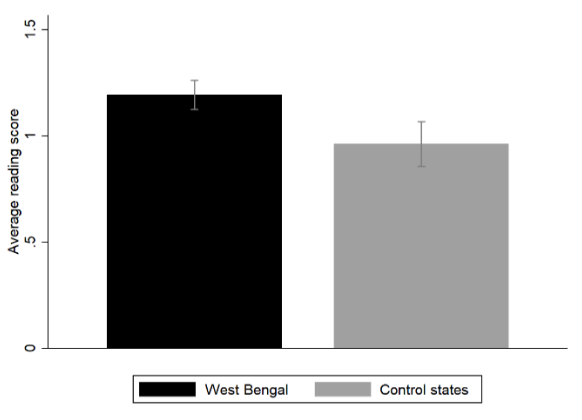
Notes: DISE data from 2010-2017 were used for schools in West Bengal only.

Figure 3: Pre-program comparison of availability of pre-primary in government schools and learning outcomes of government school pre-primary children between West Bengal and other states

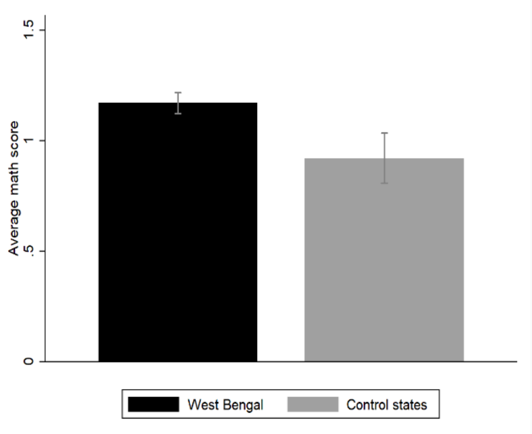
a) Percentage of government schools with pre-primary



b) Reading score of 5-6 years old

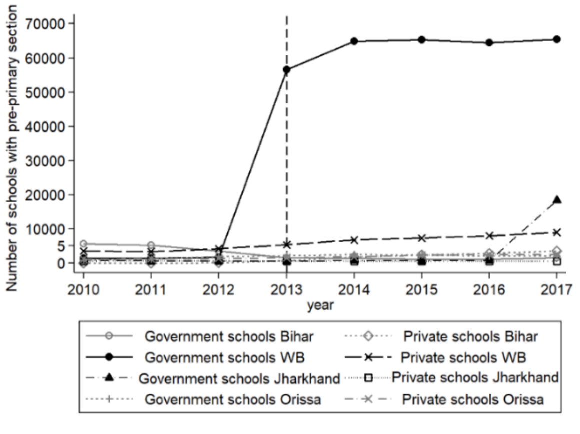


c) Math score of 5-6 years old



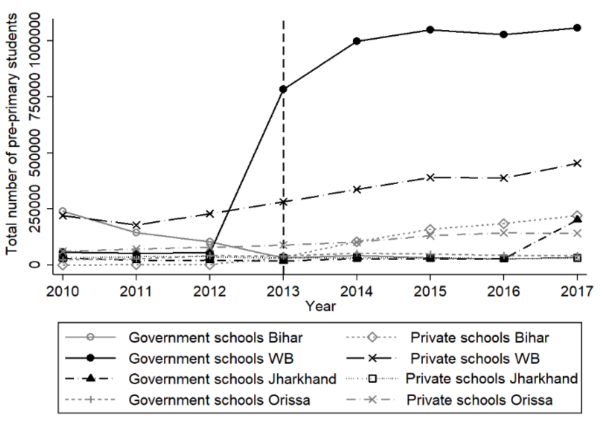
Note: Control states include Bihar, Orissa, and Jharkhand. DISE data from 2009 to 2012 were used to calculate the proportion of government schools with pre-primary. ASER data from 2009 to 2012 were used to calculate the test scores of government school children aged 5-6 years

Figure 4: Comparison of schools with pre-primary section between West Bengal & neighbouring states



Notes: DISE data from 2010-2017 were used for the states of West Bengal, Bihar, Jharkhand, and Orissa.

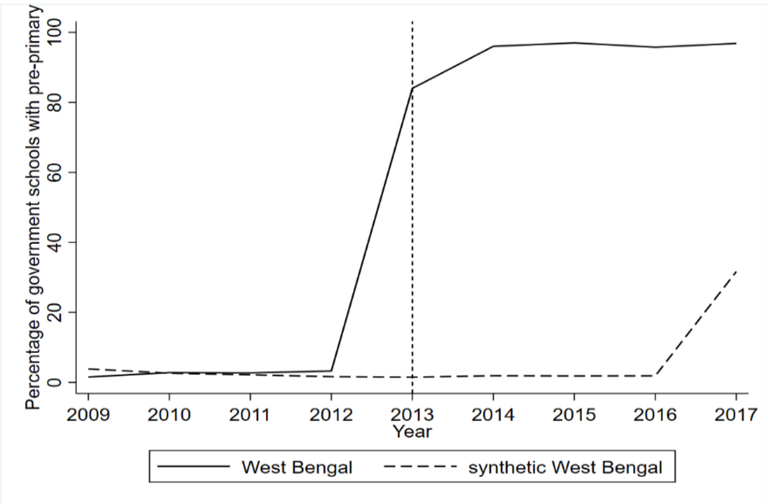
Figure 5: Comparison of student enrolment in pre-primary section between West Bengal & neighbouring states



Notes: DISE data from 2010-2017 were used for the states of West Bengal, Bihar, Jharkhand, and Orissa.

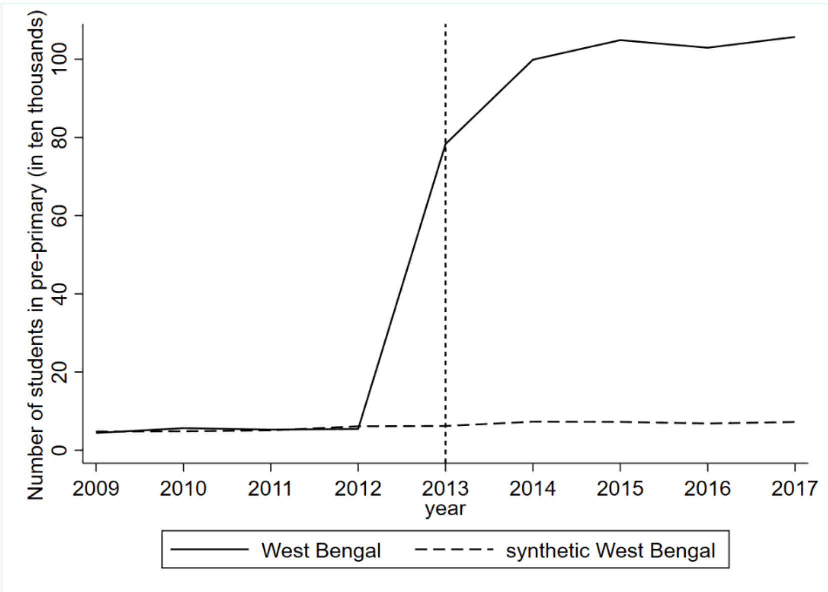
Figure 6: Comparison of pre-primary enrolment between West Bengal and synthetic control

a) Percentage of government schools with pre-primary



Note: The synthetic control method was used to generate this figure. DISE data from 2009 to 2017 were used. The predictor variables are state-wise average percentage of government schools with pre-primary (for the years 2009-2012), state-wise average number of classrooms in government schools, proportion of government schools with electricity, and playground.

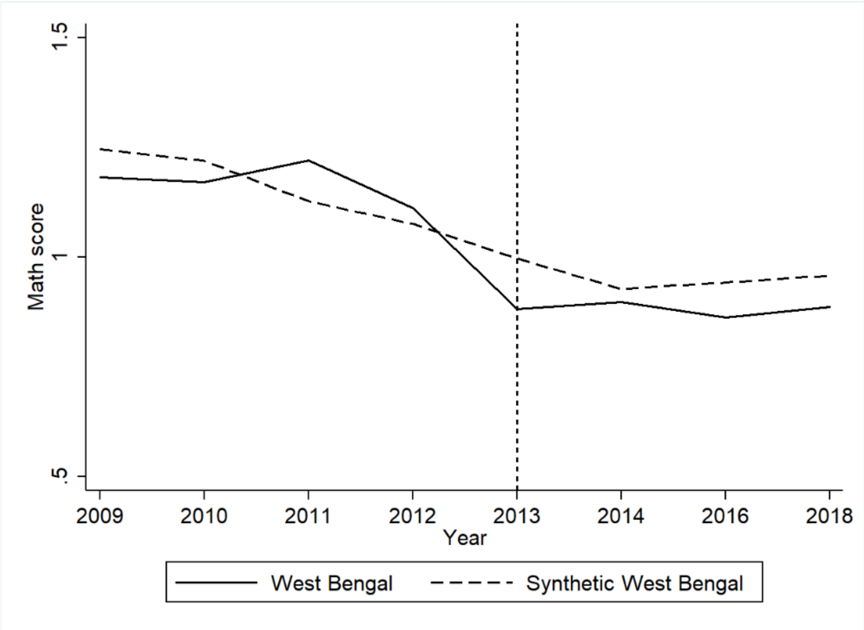
b) Number of pre-primary students in government schools



Note: The synthetic control method was used to generate this figure. DISE data from 2009 to 2017 were used. The predictor variables are state-wise total number of pre-primary students in government schools (for the years 2009-2012), state-wise average number of classrooms in government schools, proportion of government schools with electricity, and playground.

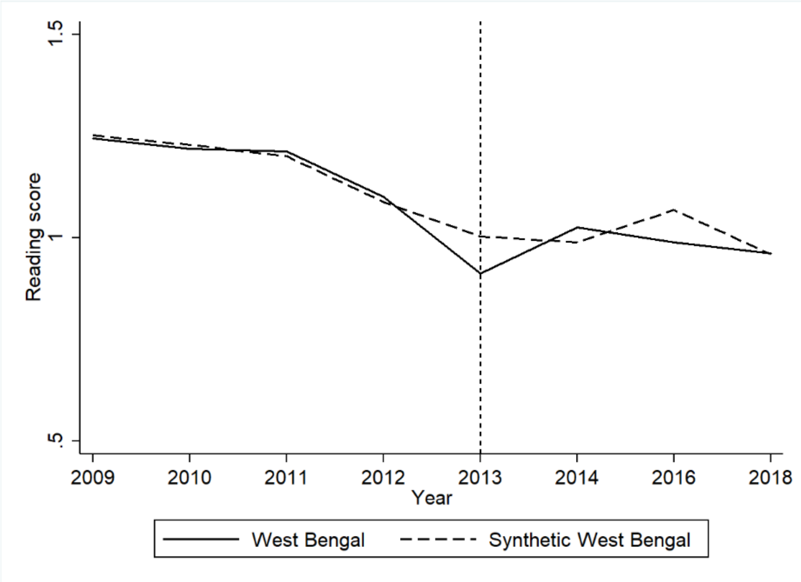
Figure 7: Learning outcomes for WB and synthetic control

7.a) Math score of 5-6 years old children from government schools



Note: The synthetic control method was used to generate this figure. Only children aged 5 and 6 were considered. ASER data from 2009 to 2018 were used. The predictor variables are average math score (2009-2012), state-wise proportion of households that are fully cemented, average household size, proportion of households having a TV, proportion of households with electricity, proportion of households with toilet, proportion of mothers who went to school at some point of time, proportion of villages with pucca road, proportion of villages with a bank, proportion of villages with post office, proportion of villages with private school and proportion of villages with government primary school.

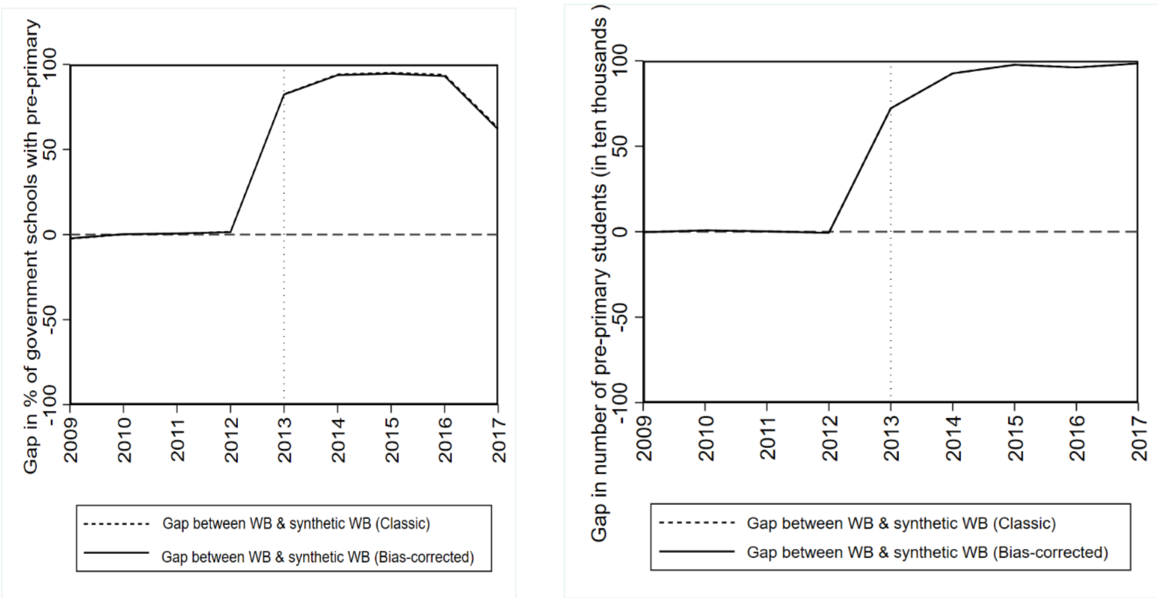
7.b) Reading score of 5-6 years old children from government schools



Note: The synthetic control method was used to generate this figure. Only children aged 5 and 6 were considered. ASER data from 2009 to 2018 were used. The predictor variables are average reading score (2009-2012), state-wise proportion of households that are fully cemented, average household size, proportion of households having a TV, proportion of households with electricity, proportion of households with toilet, proportion of mothers who went to school at some point of time, proportion of villages with pucca road, proportion of villages with a bank, proportion of villages with post office, proportion of villages with private school and proportion of villages with government primary school

Figure 8: Effect graphs for Pre-primary enrolment using synthetic control method

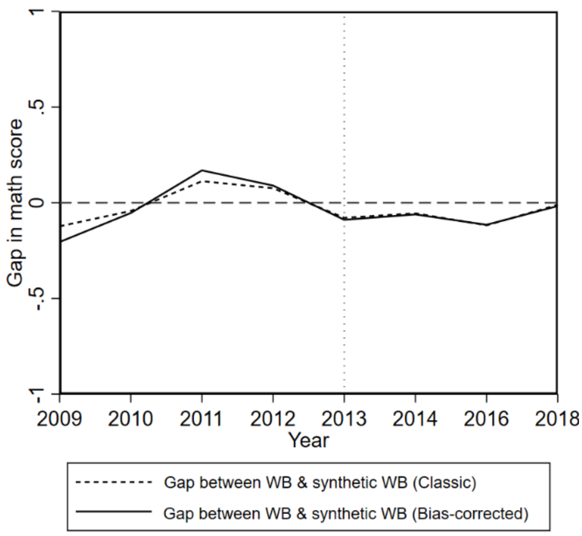
Panel A: Schools with pre-primary Panel B: Number of students in pre-primary



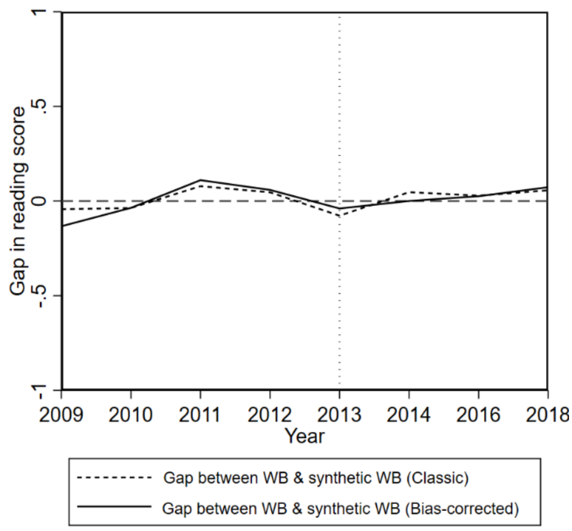
Notes: Difference in the outcome variable (number of government schools with pre-primary, number of students in pre-primary of government schools) between West Bengal and its synthetic control counterpart is referred to as 'gap'.

Figure 9: Effect graphs for learning outcome using synthetic control method

Panel A: Gap in math score



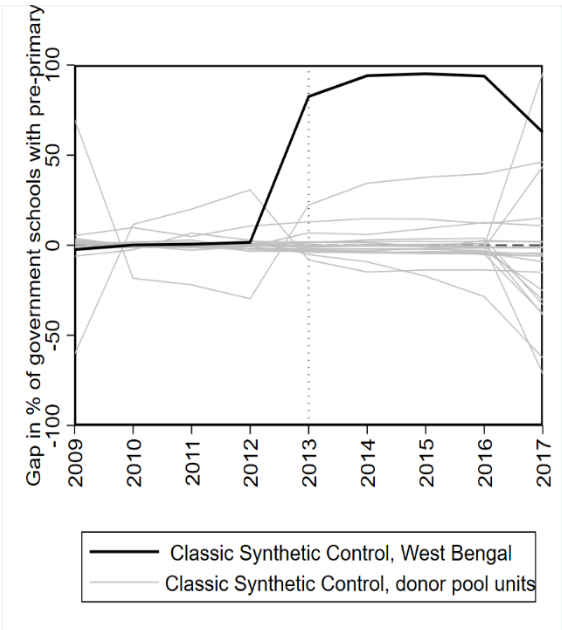
Panel B: Gap in reading score

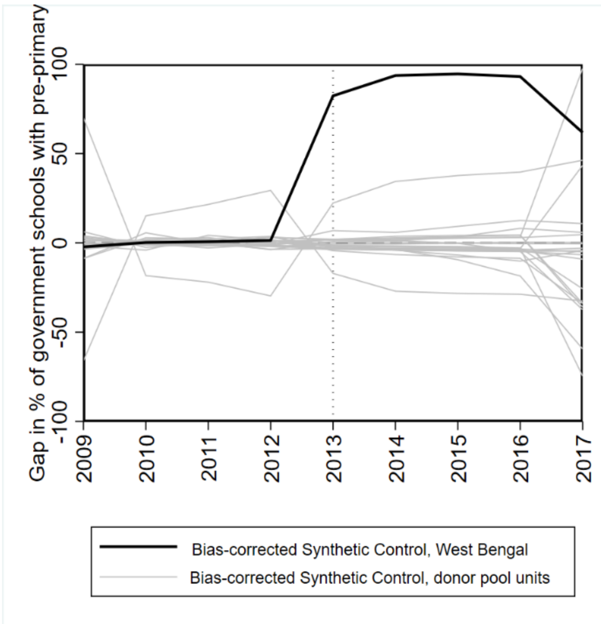


Notes: Difference in the outcome variable (math score of government school pre-primary children, reading score of government school pre-primary children) between West Bengal and its synthetic control counterpart is referred to as 'gap'.

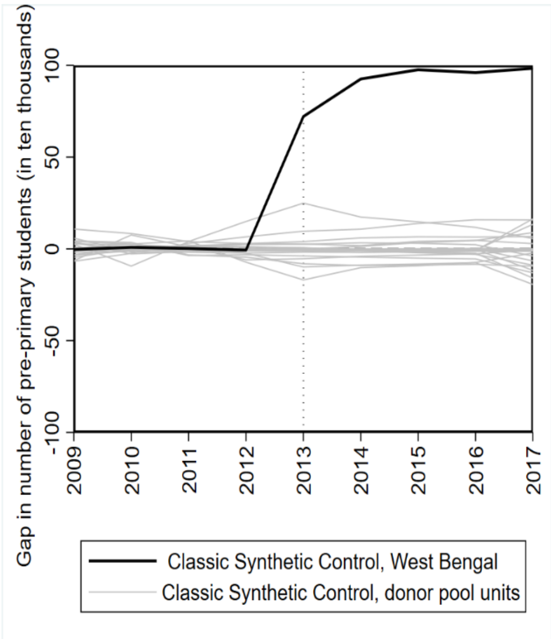
Figure 10: Effect and placebo graphs for pre-primary enrolment using synthetic control method

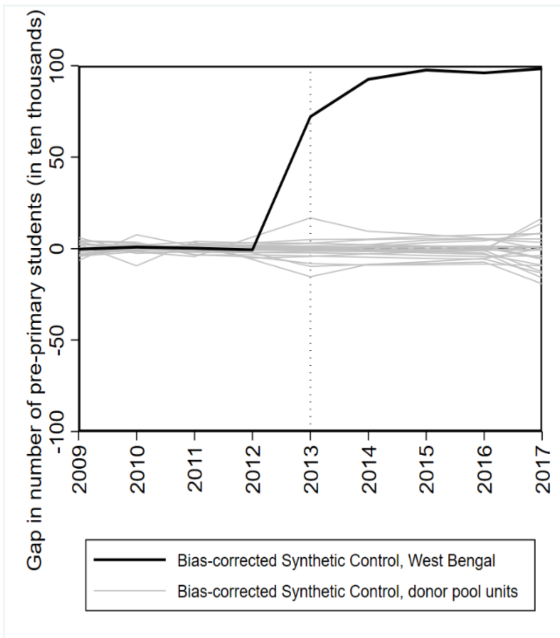
Panel A: Percentage of schools with pre-primary





Panel B: Number of students in pre-primary

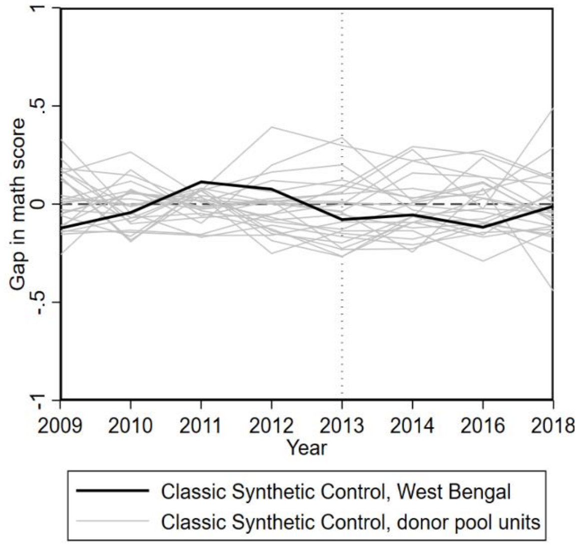


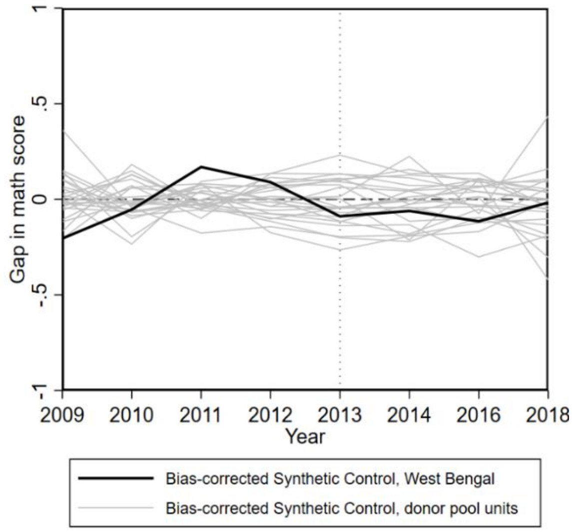


Notes: See notes to Figure 8. Same exercise has been carried out for all the states in the donor pool.

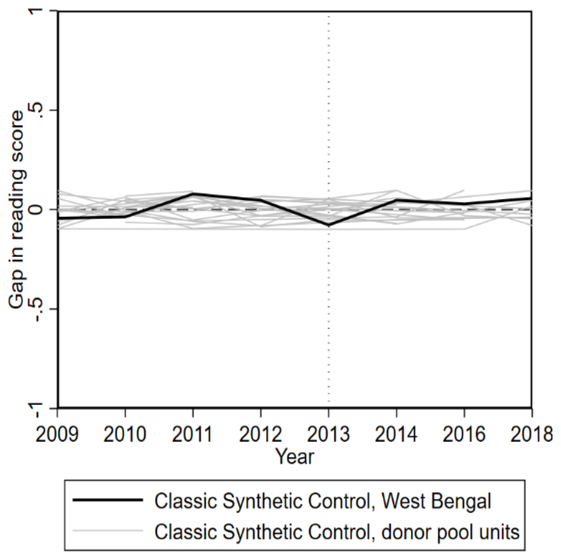
Figure 11: Effect and placebo graphs for learning outcomes using synthetic control method

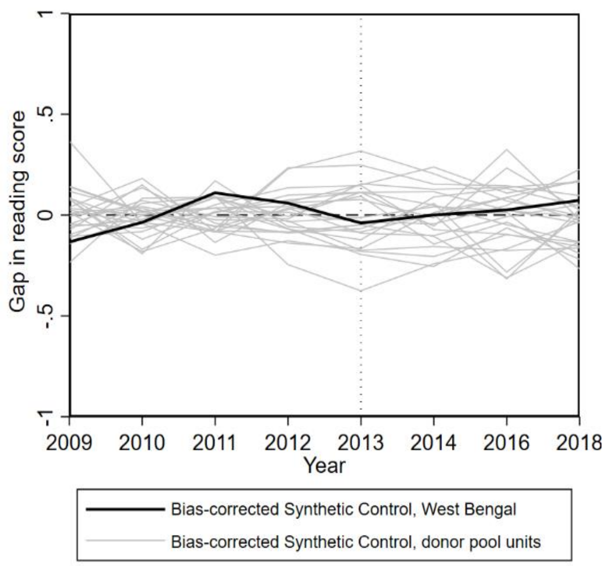
Panel A: Gap in math score





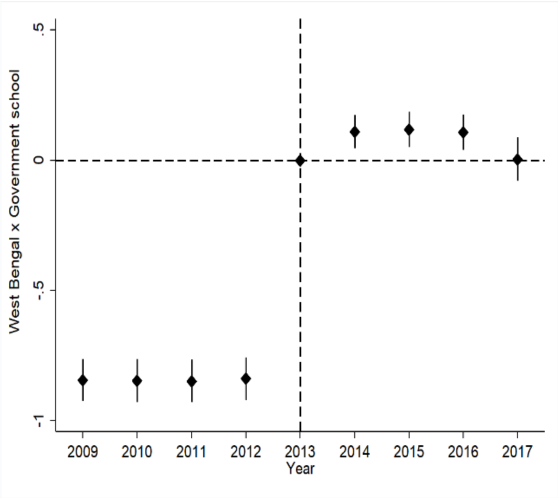
Panel B: Gap in reading score

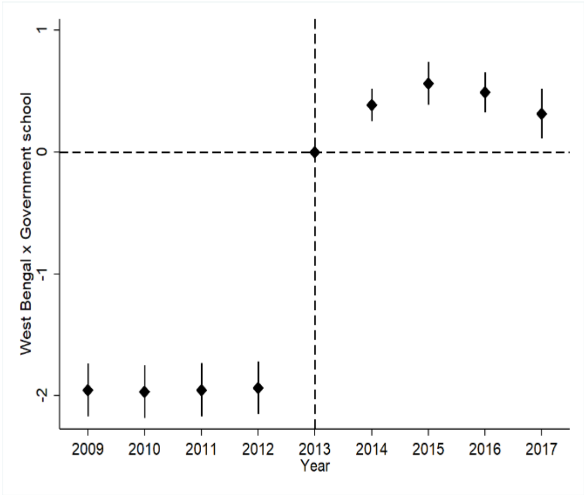




Notes: See notes to Figure 9. Same exercise has been carried out for all the states in the donor pool.

Figure 12: Event study graphs for pre-primary enrolment by year



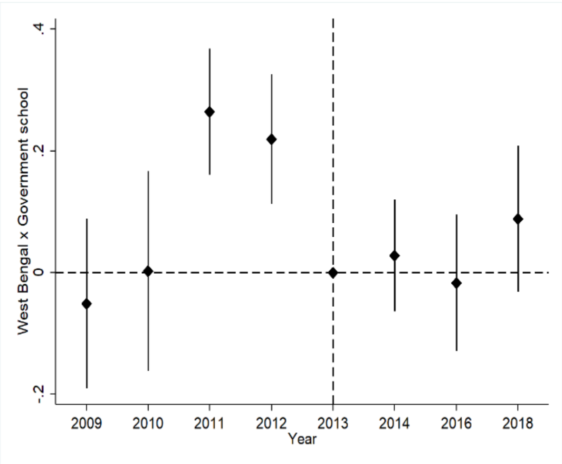


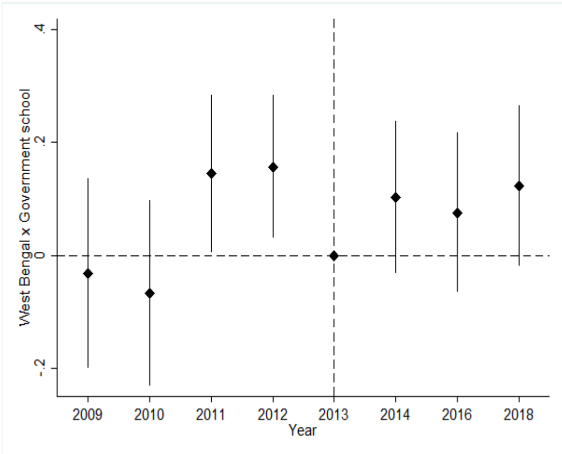
a) Availability of pre-primary

b) Student enrolment in pre-primary

Notes: DISE data from 2009-2017 were used.

Figure 13: Event study graphs for learning outcomes of pre-primary children by year





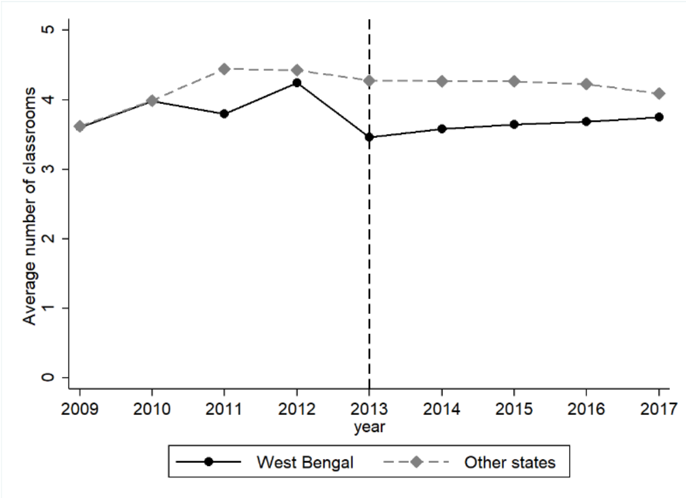
a) Math score

b) Reading score

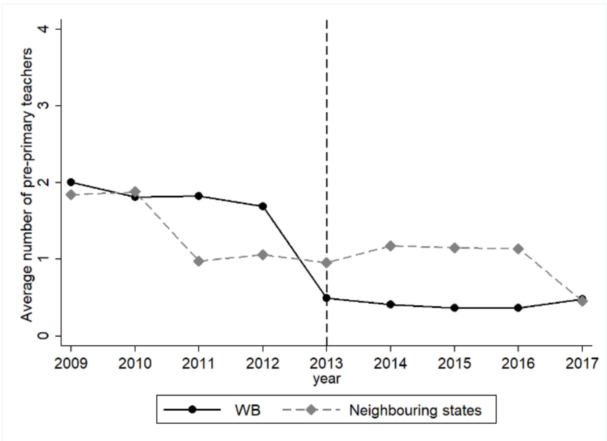
Notes: ASER data from 2009-2018 were used.

Figure 14: School infrastructure in government schools in West Bengal and neighbouring states

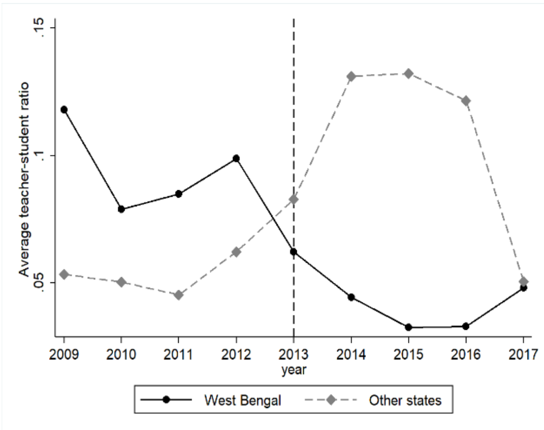
Panel A: Number of classrooms



Panel B: Average number of pre-primary teachers



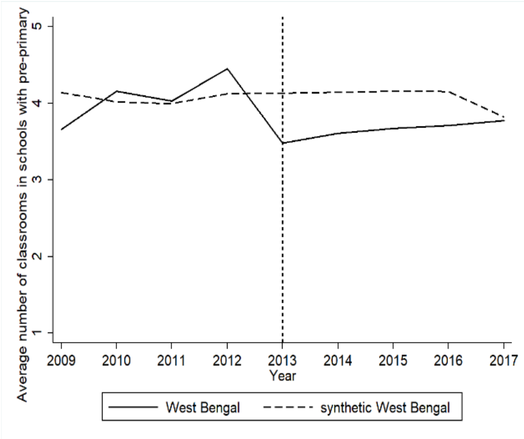
Panel C: Average student-teacher ratio



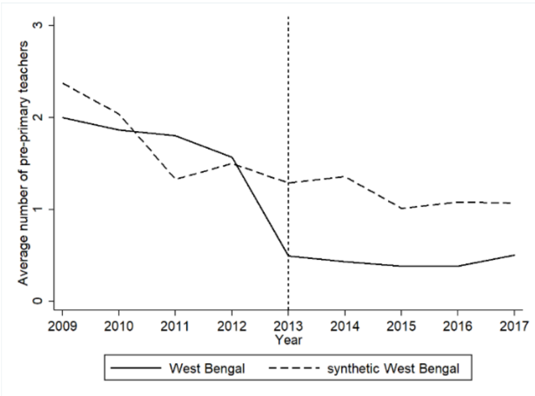
Notes: School infrastructure graphs have been created using DISE data from 2009 to 2017 for government schools of West Bengal and neighbouring states. Neighbouring states include Bihar, Jharkhand, Orissa. Only those schools with pre-primary section have been considered while creating the graphs.

Figure 15: School infrastructure in West Bengal & synthetic control states

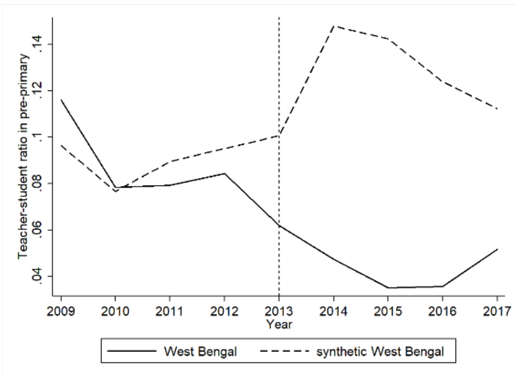
Panel A: Average number of classrooms in government schools with pre-primary



Panel B: Average number of pre-primary teachers in government schools with pre-primary



Panel B: Average teacher-student ratio in government schools with pre-primary



Note: The synthetic control method was used to generate this figure. DISE data from 2009 to 2017 were used. The predictor variables are state-wise total number of pre-primary students in government schools (for the years 2009-2012), state-wise average number of classrooms in government schools, proportion of government schools with electricity, and playground.

Tables

Table 1: DD result of availability of pre-primary and enrolment in pre-primary for schools in West Bengal

	Availability of pre-primary schools	Log (1+number of pre-primary students)
Government school	-0.39***	-1.20***
	(0.05)	(0.18)
Post	0.32**	0.52
	(0.11)	(0.42)
Government school x Post	**0.58*** **	**1.39*** **
	(0.05)	**(0.18)**
School-level controls	Yes	Yes
Year fixed effect	Yes	Yes
District fixed effect	Yes	Yes
District-specific linear trend	Yes	Yes
No of Observations	618,365	618,365
R Squared	0.78	0.57

Note: The school-level control variables include the total number of classrooms, the total number of teachers, and the availability of electricity and a playground in the school. Standard errors are clustered at the district level. Source: DISE 2009–17; authors' own calculations. ***, **, * denote significance levels at 1%, 5%, 10% respectively.

Table 2: DD result of availability of pre-primary and enrolment for government schools in West Bengal and neighbouring states

	Availability of pre-primary schools	Log (1+number of pre-primary students)
Post	-0.59***	-1.26***
	(0.14)	(0.32)
West Bengal x Post	**0.91*** **	**2.06*** **
	(0.04)	**(0.11)**
School-level controls	Yes	Yes
Year fixed effect	Yes	Yes
District fixed effect	Yes	Yes
District-specific linear trend	Yes	Yes
No of Observations	1,989,607	1,989,607
R Squared	0.75	0.65

Note: The school-level control variables include the total number of classrooms, the total number of teachers, and the availability of electricity and a playground in the school. Standard errors are clustered at the district level. Source: DISE 2009–17; authors' own calculations. ***, **, * denote significance levels at 1%, 5%, 10% respectively.

Table 3: TD result of pre-primary availability and enrolment in West Bengal and neighbouring states

	Availability of pre-primary schools	Log (1+number of pre-primary students)
Post	-0.39***	-0.75**
	(0.13)	(0.30)
Government school	-0.53***	-1.74***
	(0.02)	(0.11)
West Bengal x Post	0.18***	0.30*
	(0.05)	(0.18)
Post x Government school	-0.15***	-0.43***
	(0.02)	(0.10)
Government school x West Bengal	0.13**	0.38*
	(0.05)	(0.21)
West Bengal x Government school x Post	**0.72*** **	**1.73*** **
	(0.06)	**(0.20)**
School level controls	Yes	Yes
District-specific linear trend	Yes	Yes
No of Observations	2,110,088	2,110,088
R Squared	0.72	0.59

Notes: The school-level control variables include the total number of classrooms, the total number of teachers, and the availability of electricity and playground in the school. Standard errors are clustered at the district level. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: DISE 2009–17; authors' own calculations.

Table 4: DD result of math and reading score for pre-primary children in West Bengal

Test Score	Raw Math Score		Raw Reading Score		Age-wise standardized math score		Age-wise standardized reading score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Government school	-0.48***	-0.32***	-0.48***	-0.31***	0.03	0.23***	0.00	0.19***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
Post	-0.29	-0.37	-0.01	-0.11	-0.04	-0.06	0.07	0.04
	(0.27)	(0.25)	(0.32)	(0.32)	(0.29)	(0.29)	(0.28)	(0.31)
Government school x post	-0.05	-0.04	-0.11*	-0.10	-0.01	-0.02	-0.02	-0.03
	(0.04)	(0.04)	(0.06)	(0.07)	(0.04)	(0.05)	(0.06)	(0.07)
Household level controls	No	Yes	No	Yes	No	Yes	No	Yes
Child level controls	No	Yes	No	Yes	No	Yes	No	Yes
Village level controls	No	Yes	Yes	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-specific linear trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Observations	9,097	8,316	9,205	8,407	9,097	8,316	9,205	8,407
R Squared	0.12	0.21	0.10	0.19	0.06	0.11	0.06	0.11

Note: Household controls include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics include the child's age, whether the mother went to school. Village-level factors controlled for include whether the village has a government primary school, private school, village post office, bank, pucca road in the village. linear regression coefficients are presented with clustered standard errors at the district level given in parenthesis. Children aged 5 and 6 years have been considered for analysis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: ASER 2009–18; authors' own calculations

Table 5: DD result of math and reading score for pre-primary government school children in West Bengal and neighbouring states

Test Score	Raw Math Score		Raw Reading Score		Age-wise standardized math score		Age-wise standardized reading score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Post	-0.23**	-0.18**	-0.20**	-0.16	-0.01	0.15	0.01	0.15
	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.11)	(0.09)	(0.10)
West Bengal x post	-0.26***	-0.26***	-0.13*	-0.12**	0.01	-0.02	0.02	0.00
	(0.05)	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)
Household level controls	No	Yes	No	Yes	No	Yes	No	Yes
Child level controls	No	Yes	No	Yes	No	Yes	No	Yes
Village level controls	No	Yes	Yes	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District specific linear trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Observations	75,848	62,629	76,512	63,110	75,848	62,629	76,512	63,110
R Squared	0.09	0.13	0.10	0.14	0.05	0.13	0.05	0.08

Note: Household controls include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics include the child's age, whether the mother went to school. Village-level factors controlled for include whether the village has a government primary school, private school, village post office, bank, pucca road in the village. linear regression coefficients are presented with clustered standard errors at the district level given in parenthesis. Child aged 5 and 6 years have been considered for analysis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: ASER 2009–18; authors' own calculations

Table 6: TD result of math and reading score for pre-primary children in West Bengal and neighbouring states

Test Score	Raw Math Score		Raw Reading Score		Age-wise standardized math score		Age-wise standardized reading score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Post	0.02	0.02	0.11	0.11	-0.06	0.06	-0.04	0.06
	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Government school	-0.48***	-0.36***	-0.53***	-0.39***	0.02	0.17***	0.02	0.17***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
West Bengal x post	-0.26***	-0.27***	-0.13	-0.13	0.05	0.03	0.07	0.06
	(0.06)	(0.06)	(0.08)	(0.09)	(0.06)	(0.06)	(0.08)	(0.09)
Post x Government school	-0.12***	-0.10***	-0.14***	-0.12***	0.03	0.04	0.02	0.04
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Government school x West Bengal	0.02	0.01	0.06	0.05	0.01	0.00	-0.02	-0.03
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
West Bengal x Government School x post	0.07	0.06	0.03	0.02	-0.04	-0.06	-0.05	-0.07
	(0.05)	(0.05)	(0.07)	(0.08)	(0.05)	(0.05)	(0.07)	(0.07)
Household level controls	No	Yes	No	Yes	No	Yes	No	Yes
Child level controls	No	Yes	No	Yes	No	Yes	No	Yes
Village level controls	No	Yes	Yes	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District specific linear trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Observations	87,412	72,517	88,215	73,058	87,472	72,517	88,215	73,058
R Squared	0.11	0.16	0.12	0.17	0.04	0.07	0.04	0.07

Note: Household controls include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics include the child's age, whether the mother went to school. Village-level factors controlled for include whether the village has a government primary school, private school, village post office, bank, pucca road in the village. linear regression coefficients are presented with clustered standard errors at the district level given in parenthesis. Child aged 5 and 6 years have been considered for analysis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: ASER 2009–18; authors' own calculations

Table 7: DD result of school infrastructure for government schools with pre-primary in West Bengal and neighbouring states

	Log(1+Total number of classrooms)	Log(1+number of pre-primary teachers)	Teacher-student ratio
Post	0.56***	-0.57***	-0.07***
	(0.08)	(0.11)	(0.03)
West Bengal x Post	**-0.17*** **	**-0.75*** **	**-0.05*** **
	(0.04)	**(0.07)**	**(0.02)**
School-level controls	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
District fixed effect	Yes	Yes	Yes
District-specific linear trend	Yes	Yes	Yes
No of Observations	459,214	457,837	408,602
R Squared	0.39	0.34	0.04

Note: The school-level control variables for the second regression include the total number of classrooms, and the availability of electricity and playground in the school. In the first regression, since the dependent variable is total number of classrooms itself, it has been excluded from the list of control variables. In the third regression, Teacher-student ratio has been calculated. The corresponding control variables are same as in the second regression. The Standard errors are clustered at the district level. Only schools with pre-primary section have been considered. Source: DISE 2009–17; authors' own calculations

Appendix

Appendix Table 1: Parallel trend assumption on availability of pre-primary schools and pre-primary enrolment for West Bengal

	Availability of pre-primary schools	Log (1+number of pre-primary students)
Government school	-0.30***	-1.15***
	(0.04)	(0.17)
Time	0.08**	0.14
	(0.04)	(0.16)
Government school x Post	-0.10**	-0.22
	(0.04)	(0.15)
School-level controls	Yes	Yes
Year fixed effect	Yes	Yes
District-specific linear trend	Yes	Yes
No of Observations	244,407	244,407
R Squared	0.33	0.28

Note: The school-level control variables include the total number of classrooms, the total number of teachers, and the availability of electricity and playground in the school. Standard errors are clustered at the district level. Source: DISE 2009–12; authors' own calculations

Appendix Table 2: DD result on availability of pre-primary schools and pre-primary enrolment for government schools in West Bengal and neighbouring states

	Availability of pre-primary schools	Log (1+number of pre-primary students)
Post	0.02***	0.01
	(0.01)	(0.01)
West Bengal x Post	-0.01	-0.01
	(0.01)	(0.02)
School-level controls	Yes	Yes
Year fixed effect	Yes	Yes
District fixed effect	Yes	Yes
District specific linear trend	Yes	Yes
No of Observations	847,868	847,868
R Squared	0.04	0.03

Note: The school-level control variables include the total number of classrooms, the total number of teachers, and availability of electricity and playground in the school. Standard errors are clustered at the district level. Source: DISE 2009–12; authors' own calculations. ***, **, * denote significance levels at 1%, 5%, 10% respectively.

Appendix Table 3: Parallel trend assumption on availability and enrolment in pre-primary between West Bengal & neighbouring states (Triple difference)

	Availability of pre-primary schools	Log (1+number of pre-primary students)
Time	0.14***	0.37***
	(0.03)	(0.10)
Government school	-0.41***	-1.49***
	(0.04)	(0.15)
West Bengal x Time	-0.03	-0.15
	(0.05)	(0.17)
Time x Government school	-0.12***	-0.37***
	(0.03)	(0.10)
Government school x West Bengal	0.09	0.26
	(0.06)	(0.22)
West Bengal x Government school x Time	0.02	0.15
	(0.05)	(0.18)
School level controls	Yes	Yes
District specific linear trend	Yes	Yes
No of Observations	890,775	890,775
R Squared	0.20	0.19

Notes: The school-level control variables include the total number of classrooms, total number of teachers and availability of electricity and playground in the school. Linear regression coefficients are presented with clustered standard errors at the state level given in parenthesis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: DISE 2009–12; authors' own calculations.

Appendix Table 4:

Testing parallel trend assumption on learning outcome for government and private school pre-primary children of West Bengal

Test Score	Raw Math Score		Raw Reading Score		Age-wise standardized math score		Age-wise standardized reading score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Government school	-0.62***	-0.47***	-0.53***	-0.34***	-0.01	0.19**	-0.03	0.17*
	(0.11)	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)
Post	-0.10	-0.11	-0.10	-0.06	-0.13	-0.17	-0.16	-0.17
	(0.16)	(0.18)	(0.18)	(0.20)	(0.17)	(0.19)	(0.17)	(0.19)
Government school x post	0.21*	0.23**	0.09	0.08	0.06	0.07	0.07	0.06
	(0.12)	(0.11)	(0.13)	(0.13)	(0.11)	(0.02)	(0.11)	(0.12)
Household level controls	No	Yes	No	Yes	No	Yes	No	Yes
Child level controls	No	Yes	No	Yes	No	Yes	No	Yes
Village level controls	No	Yes	Yes	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District specific linear trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Observations	4,625	4,096	4,709	4,164	4,625	4,096	4,709	4,164
R Squared	0.11	0.19	0.11	0.19	0.07	0.13	0.07	0.14

Note: Household controls include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics include the child's age, whether the mother went to school. Village-level factors controlled for include whether the village has a government primary school, private school, village post office, bank, pucca road in the village. linear regression coefficients are presented with clustered standard errors at the district level given in parenthesis. Child aged 5 and 6 years have been considered for analysis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: ASER 2009–12; authors' own calculations

Appendix Table 5: Testing parallel trend assumption on learning outcome for government school pre-primary children of West Bengal and neighbouring states

Test Score	Raw Math Score		Raw Reading Score		Age-wise standardized math score		Age-wise standardized reading score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Post	-0.22***	-0.21***	-0.09	-0.05	0.01	0.00	0.01	0.03
	(0.07)	(0.07)	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.08)
West Bengal x post	-0.05	-0.01	-0.12	-0.12	-0.02	0.02	-0.03	-0.02
	(0.10)	(0.11)	(0.12)	(0.12)	(0.10)	(0.11)	(0.11)	(0.12)
Household level controls	No	Yes	No	Yes	No	Yes	No	Yes
Child level controls	No	Yes	No	Yes	No	Yes	No	Yes
Village level controls	No	Yes	Yes	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District specific linear trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Observations	42,316	33,287	42,854	33,665	42,316	33,287	42,854	33,665
R Squared	0.10	0.13	0.09	0.13	0.07	0.10	0.07	0.10

Note: Household controls include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics include the child's age, whether the mother went to school. Village-level factors controlled for include whether the village has a government primary school, private school, village post office, bank, pucca road in the village. linear regression coefficients are presented with clustered standard errors at the district level given in parenthesis. Child aged 5 and 6 years have been considered for analysis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: ASER 2009–12; authors' own calculations

Appendix Table 6: Testing parallel trend assumption on learning outcomes for both government and private school pre-primary children of West Bengal and control states (Triple difference)

Test Score	Raw Math Score		Raw Reading Score		Age-wise standardized math score		Age-wise standardized reading score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Post	-0.16**	-0.10	0.11	0.05	-0.04	-0.02	-0.04	-0.03
	(0.08)	(0.08)	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Government school	-0.43***	-0.28***	-0.41***	-0.29***	-0.02	0.14***	-0.01	0.11**
	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)	(0.05)
West Bengal x post	-0.33**	-0.35**	-0.38***	-0.36**	-0.03	-0.03	-0.04	-0.02
	(0.14)	(0.14)	(0.14)	(0.15)	(0.13)	(0.14)	(0.13)	(0.14)
Post x Government school	-0.07	-0.10**	-0.16***	-0.13**	0.05	0.03	0.05	0.07
	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)	(0.05)	(0.04)	(0.05)
Government school x West Bengal	-0.19*	-0.23**	-0.11	-0.10	0.01	-0.02	-0.02	0.00
	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)	(0.11)
West Bengal x Government School x post	0.29**	0.33***	0.25*	0.21	0.02	0.04	0.02	-0.01
	(0.13)	(0.12)	(0.14)	(0.14)	(0.12)	(0.11)	(0.12)	(0.13)
Household level controls	No	Yes	No	Yes	No	Yes	No	Yes
Child level controls	No	Yes	No	Yes	No	Yes	No	Yes
Village level controls	No	Yes	Yes	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District specific linear trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Observations	47,035	37,022	47,635	37,448	47,035	37,022	47,635	37,448
R Squared	0.11	0.15	0.11	0.15	0.07	0.09	0.06	0.09

Note: Household controls include the number of household members, whether the house was pucca or not; whether it had electricity; possession of a TV, presence of a toilet. Child-level characteristics include the child's age, whether the mother went to school. Village-level factors controlled for include whether the village has a government primary school, private school, village post office, bank, pucca road in the village. linear regression coefficients are presented with clustered standard errors at the district level given in parenthesis. Child aged 5 and 6 years have been considered for analysis. ***=significant at 1% level, **=significant at 5% level, *=significant at 10% level. Source: ASER 2009–12; authors' own calculations

Appendix Table 1: Weights and predictor balance for availability of pre-primary corresponding to Figure 6 a)

Panel A: Weights for the states

State name	Weight
Himachal Pradesh	0
Punjab	0
Uttarakhand	0
Haryana	0
Rajasthan	0
Uttar Pradesh	0
Bihar	0
Sikkim	0
Nagaland	0
Manipur	0
Tripura	0.22
Meghalaya	0
Assam	0
Jharkhand	0.76
Orissa	0
Chhattisgarh	0
Madhya Pradesh	0
Gujarat	0
Maharashtra	0
Karnataka	0
Kerala	0.02
Tamil Nādu	0

Panel B: Predictor balance

	Treated	Synthetic
Percentage of schools with pre-primary	2.58	2.75
Proportion of government schools with electricity	0.32	0.10
Proportion of government schools with playground	0.36	0.36
Average number of classrooms in government schools	4.34	4.25

All predictor variables are averaged over the period 2009 to 2012

Appendix Table 2: Weights and predictor balance for number of pre-primary students corresponding to Figure 6 b)

Panel A: Weights for the states

State name	Weight
Himachal Pradesh	0
Punjab	0.01
Uttarakhand	0
Haryana	0.01
Rajasthan	0.01
Uttar Pradesh	0
Bihar	0.01
Sikkim	0.04
Nagaland	0.02
Manipur	0
Tripura	0
Meghalaya	0.01
Assam	0
Jharkhand	0.02
Orissa	0.69
Chhattisgarh	0.01
Madhya Pradesh	0.01
Gujarat	0.01
Maharashtra	0.01
Karnataka	0.02
Kerala	0.11
Tamilnadu	0.01

Panel B: Predictor balance

	Treated	Synthetic
Number of students in pre-primary	51974.25	51993.06
Proportion of government schools with electricity	0.32	0.32
Proportion of government schools with playground	0.36	0.36
Average number of classrooms in government schools	4.34	4.34

All predictor variables are averaged over the period 2009 to 2012

Appendix Table 3: Weights and predictor balance for math score corresponding to Figure 7 a)

Panel A: Weights for the states

State name	Weight
Himachal Pradesh	0
Punjab	0
Uttarakhand	0
Haryana	0
Rajasthan	0
Uttar Pradesh	0
Bihar	0.15
Sikkim	0
Nagaland	0.34
Manipur	0
Mizoram	0
Tripura	0
Meghalaya	0
Assam	0.38
Jharkhand	0
Orissa	0
Chhattisgarh	0
Madhya Pradesh	0
Gujarat	0
Maharashtra	0
Andhra Pradesh	0
Karnataka	0
Kerala	0.13
Tamil Nādu	0

Panel B: Predictor Balance

	Treated	Synthetic
Math score	1.17	1.17
Proportion of pucca households	0.14	0.14
Household size	6.12	6.10
Proportion of households with toilet	0.39	0.55
Proportion of household with electricity	0.58	0.65
Proportion of household with TV	0.31	0.34
Proportion of villages with pucca road	0.46	0.49
Proportion of villages with post office	0.36	0.36
Proportion of villages with bank	0.19	0.19
Proportion of villages with private school	0.28	0.41
Proportion of villages with government primary school	0.95	0.95
Proportion of household where mother went to school	0.59	0.68

All predictor variables are averaged over the period 2009 to 2012

Appendix Table 4: Weights and predictor balance for reading score corresponding to Figure 7 b)

Panel A: Weights for the states

State name	Weight
Himachal Pradesh	0.37
Punjab	0
Uttarakhand	0
Haryana	0
Rajasthan	0
Uttar Pradesh	0
Bihar	0
Sikkim	0
Nagaland	0.21
Manipur	0.06
Mizoram	0
Tripura	0
Meghalaya	0
Assam	0.06
Jharkhand	0.24
Orissa	0
Chhattisgarh	0
Madhya Pradesh	0
Gujarat	0
Maharashtra	0
Andhra Pradesh	0
Karnataka	0
Kerala	0.06
Tamil Nādu	0

Panel B: Predictor Balance

	Treated	Synthetic
Reading score	1.19	1.19
Proportion of pucca households	0.14	0.23
Household size	6.12	6.17
Proportion of households with toilet	0.39	0.56
Proportion of household with electricity	0.58	0.82
Proportion of household with TV	0.31	0.51
Proportion of villages with pucca road	0.46	0.49
Proportion of villages with post office	0.36	0.35
Proportion of villages with bank	0.19	0.17
Proportion of villages with private school	0.28	0.33
Proportion of villages with government primary school	0.95	0.90
Proportion of household where mother went to school	0.59	0.68

All predictor variables are averaged over the period 2009 to 2012