Supporting Learning In and Out of School: Experimental Evidence from India

Martina Björkman Nyqvist[1] Andrea Guariso[2]

6 January 2021

Abstract: In this paper we study an innovative education program designed to address two challenges faced by many primary school children in developing countries: 1) the gap between their actual knowledge and the level targeted by teachers in class, 2) the lack of support for their studying and learning activities outside of school hours. The program, implemented in India, combines an *in-school* pedagogical intervention with the creation of *out-of-school* study groups. We designed a randomized experiment with factorial design to assess the effectiveness of the full program as well as of its individual *in-school* and *out-of-school* components. Results show that the full program significantly increases test scores in both mathematics and language. However, when implemented independently, the two components have no impact. The analysis reveals the importance of the timing and the intensity of the programs and finds evidence of inputs substitution taking place in the schools. Overall, the findings indicate that learning programs are more likely to succeed when they adopt a multidimensional approach and strengthen the children's learning process both in and out of school.

Keywords: Education policy; Learning Outcomes; Community engagement, In-school pedagogical program; India; Randomized Controlled Trial

JEL codes: C93; I21; O15

We thank Abhijeet Singh, Timothée Demont, and seminar participants at Trinity College Dublin and Aix-Marseille School of Economics (AMSE) for helpful comments and suggestions, and Merve Demirel, Artur Obminski, and Harry Humes for invaluable research assistance. We also thank J-PAL South Asia and its staff, specifically Srijana Chandrashekhar; Mrignaina Tikku; Dwithiya Raghavan; Tanima; Akanksha Saletore; Maaike Bijker; Rithika Nair and Shobini Mukerji for support and management of the data collection and fieldwork. Finally, we thank Rukmini Banerjee and Saveri Kulshreshth at Pratham India and Ingrid Eelde Koivisto at Pratham Sweden for insightful discussions about the programs and for the constant interactions throughout the study, as well as for sharing administrative data. Financial support from Carl Bennet AB, the Swedish Research Council (2016-05615), J-PAL CaTCh Initiative and the Mistra Foundation is greatly appreciated. Andrea also acknowledges funding from the TCD Arts and Social Sciences Benefactions (ASSB) Fund. The study received ethical approval from IFMR Human Subjects Committee (IRB00007107) and TCD Ethic Committee (05062018).

[1] Stockholm School of Economics and Misum

[2] Trinity College Dublin *(corresponding author,* guarisoa@tcd.ie*)*

1. Introduction

Developing countries have achieved dramatic gains in school enrollment over the past two decades, pushing the average net enrollment rate in primary schools above 90% (World Bank, 2018). This success followed the decision to make education top priority for the global development agenda (United Nations, 2015), which led to large increases in government expenditure on education (World Bank, 2017). Yet, learning levels remain stubbornly low and increases in quantity of schooling has not directly translated into higher learning (e.g. Glewwe and Muralidharan, 2016; Altinok et al., 2018). India is a case in point: according to the Annual Status of Education Report (ASER, 2018) more than 96% of all children in the age group of 6-14 years are now enrolled in school, but the share of children in grade 5 that can read a grade 2 level text or solve a grade 2 subtraction problems remains well below 50%, with very little progress over the past decade. There is an urgent need to better understand the constraints hampering children's learning achievements and to develop innovative programs that can relax them.

There are in particular two challenges related to children's learning and studying practices faced by many primary school children across the globe, that appear particularly severe in developing countries settings. The first one is that the level of classroom instructions is often too advanced for the students' actual knowledge level. This stems from educational systems that tend to put more emphasis on completing the planned curriculum rather than making sure everybody understands and reaches minimum standards, while at the same time automatically promoting students to the next grades (Banerjee and Duflo, 2012; Muralidharan, et al, 2019). The second challenge relates to the scarcity of opportunities and support for students to study and learn outside of school hours. This can be partly explained by the limited resources available in many settings, but is also influenced by the widespread view that educational learning is responsibility of the school and therefore confined within regular class time (Banerji and Chavan, 2016).

This paper studies a program specifically designed to address these two challenges. We partnered with the largest NGO working in the Indian education sector (Pratham) to study an innovative primary education program. The program has two key components: an *in-school* pedagogical intervention, which is directly implemented by the NGO, and an *out-of-school* initiative, which is set up by the NGO but managed by the local community. The former pedagogical component is implemented in school through repeated short sessions, called *Learning Camps,* which are structured around the "Teaching at The Right Level" approach pioneered by Pratham. The simple yet powerful idea is to rearrange children across different classes based on their actual knowledge rather than on age and grade, so that teaching can be tailored to the appropriate level of knowledge. The latter component is a community engagement initiative implemented in the village, which consists in the creation of community-based *Study Groups* through which families and communities can support the learning of their children outside of school hours, while maintaining a limited time commitment. Each study group is managed by a local volunteer and brings together about seven children of different ages and grades to study together and learn from each other.

In order to study the impact of the overall program as well as of its individual components, we designed a randomized controlled trial with factorial design, across a sample of 200 villages in the state of Assam, where the NGO was expanding its activities. We randomly allocated villages to four groups: 50 hosted the full program, with Learning Camps taking place in the local primary public school and community-based Study Groups organized in the village; 50 hosted only the Learning Camps in the local primary public school; 50 hosted only the Study Groups in the villages; and 50 were assigned to the control group where there was no NGO activity during the study period. We collected baseline data in summer 2018, when we surveyed all head teachers in the primary public schools located in the study villages (N=200), a representative sample of children enrolled in grades 1 to 4 within each school (N=5,726), and a representative sample of their caregivers (N=4,592). The endline data collection took place at the end of 2019, 16 months after the start of program implementation in the intervention villages.

We report four sets of results. First, the full education program, with its combination of Learning Camps and Study Groups, is overall successful in improving children's learning. Children score on average 0.09-0.12 standard deviation higher in math and language compared to children in the control group, and the impact is statistically significant at the 5% level. These effects translate in a significant increase in the share of children that are able to achieve minimum standards in the two subjects. The share of children in grade 2 to 5 that can solve grade-2 math (i.e. subtraction) increased by 6 percentage points (20%) compared to the control group, where only 30% of the children could attain such level. Similarly, the share of children that reached grade-2 language skills (i.e. read a short story) increased by 4 percentage points (13%) compared to the control group, where again only 30% of the children could attain that level.

Second, when implemented alone, neither of the two program components had any significant impact on children's learning levels. The point estimates of the impact on both math and language tests are close to zero and we can rule out that they are similar to the estimates of the full program at the 10% significance level.

Third, we provide evidence of two factors that might have attenuated the effectiveness of the programs: limited participation in the Study Groups and inputs substitution in the schools that hosted the Learning Camps. On the former, we find that although children in the study group villages report studying more in groups after school, direct participation in the groups remained low. On the latter, we find that schools that hosted the camps invested less in physical capital over the study period and decreased their level of engagement with stakeholders (government and management committees).

Fourth, by combining our dataset with administrative information recorded by Pratham, we find that both the timing and the intensity of the program play an important role. On the timing, we exploit random variation in the scheduling of the learning camps and find that children that were exposed to the camps early on had significantly higher test scores by the time of our endline data collection, suggesting that the learning gains from the camps put them on a different learning trajectory, which enabled them to cumulate more and more learning over time. We also find that the program is relatively more effective for children that were more exposed to the program – i.e. children that attended more Learning Camps days and that lived in villages with a higher density of Study Groups. These findings are supported by an instrumental variable analysis, where we use the random allocation to treatment arms to try to get around the likely endogeneity of program exposure.

Finally, we perform a cost-effectiveness analysis and estimate a cost between 980 INR (13.7 US$) and 1,306 INR (18.3 US$) per student to raise learning outcomes by 0.1 SD. These estimates would put this program right in the middle of a list of 27 education programs aimed at improving learning outcomes that were evaluated using randomized controlled trials and for which J-PAL estimated cost effectiveness (Bhula et al, 2013). They are also in line with what other studies have found for successful pedagogical intervention in South America.

The primary contribution of this paper is to experimentally study an education program that emphasizes *both* in-school productivity and out-of-school community engagement in order to increase children's learning levels. To our knowledge, this is the first paper to rigorously assess a program that explores the complementarities between what happens in school and what happens after school. Most of the previous attention of researchers (and policy makers) has focused on the activities happening within schools to identify key constraint for the low and slow improvements in students' learning (see Glewwe and Muralidharan (2016), and Muralidharan (2017) for a review). The existing body of evidence indicates targeted pedagogical interventions as the most promising approach to improve children's learnings, especially when compared to more standard alternatives, such as providing more physical inputs to the schools (Glewwe et al., 2009; Banerjee et al., 2016). The literature has so far placed significant less emphasis on children's learning experience outside of school hours and, in particular, on how to boost family and community support to children's learning process.[1] This might appear surprising, since children spend most of their day outside school and regular practice and exercises are key for the

[1] There are few studies that look at the impact of after-school remedial education (Lakshminarayana et al, 2013; Banerjee et al, 2016), in some cases also with the support of computer-assisted learning (Linden, 2008; Lai et al, 2015; Muralidharan et al, 2019). However, most of these programs are implemented in-school right after (or before) official school hours and are often led by teachers or specifically trained personnel. Muralidharan et. al. (2019) represents a relevant exception, although it relies on an intense technology-aided instructional program implemented in specific learning centers. Romero et al (2018) is instead the only study we are aware of that evaluates cross-age tutoring, where tutors are students in the same school as tutees (although also in this case tutoring takes place in school, right after the end of the regular school day).

learning process. One constraint to parental engagement highlighted in the literature is that parents themselves might feel inadequate to provide the necessary support to their children. However, recent work by Banerji et al. (2017) shows that parental education programs do not translate into higher learning outcome of the children, suggesting that other constraints are at play.[2] One of these is likely to be the lack of time and commitment. Families and community members, burdened with many tasks competing for their attention, might find little time to support their children's learning, also (mis-)guided by the idea that educational learning is anyway responsibility of the school (Banerji and Chavan, 2016). How to create opportunities for students to learn after school and the impact this might have on the learning outcomes thus remains an open question. The community-based Study Groups considered in this paper provide a simple opportunity for children to meet and study together after school, while keeping the required time engagement from families and community members low. By studying the combination of the Study Groups with a more standard in-school pedagogical intervention, we are able to investigate potential synergies between the two components.

This paper also contributes to the literature on the replicability of educational programs. One version of the Learning Camps had indeed already been studied in 2014 in the state of Uttar Pradesh, where it was found to lead to very large learning gains (0.6-0.7 standard deviations in mathematics and language) for children in grades 3 to 5 (Banerjee et al, 2016). Contrary to these earlier findings, our results show that the learning camps alone are not effective in raising learning outcomes in Assam. This result indicates that such pedagogical intervention might be less effective in settings where students start from relatively higher learning levels, and where classes are on average smaller.[3] There are however also some important differences between the programs considered in the two studies. The version of the learning camps studied here is indeed significantly less intense, as it lasts for a total of 32 days (30 days of camps, plus 2 introduction days before the start of the school term), as opposed to 50 days in the Utter Pradesh study. Moreover, the current version has been re-designed to target all primary school grades, while the previous version only focused on children in grades 3 to 5. Our results show that in Assam the new learning camps model only works in combination with the out-of-school study groups intervention.

[2] Another study that looks at the role of parental education is Aoki (2005), who considers both the outcomes for adults and their children of a national literacy program in Ghana, finding in this case an increase in school attendance and an increase in households that report closely supporting their children's learning at home. Ganimian and Murnane (2016), in their review of the literature point out that effects of parental involvement seem to primarily occur for interventions targeted at early childhood.

[3] The average difference in (comparable) tests scores between students in the *same* grade in our study and in the Uttar Pradesh study is almost twice the average difference in test scores between two consecutive grades within the Uttar Pradesh sample, i.e. students in our sample would be on average almost two grades ahead of the students included in the Uttar Pradesh study. Moreover, average class size in our study is about half the size of the average class in the Uttar Pradesh study.

Our results are relevant to the ongoing debate on effective strategies to improve children's learning levels and highlight the role that after-school activities can play in enhancing learning programs and improving learning outcomes. This finding supports the efforts undertaken in recent years by organizations like Pratham to broaden their educational interventions to include some forms of parental and community engagement in after-school activities.

The rest of the paper is organized as follows. Section 2 briefly describes the setting and the intervention. Section 3 describes research design and data. Section 4 presents our results. Section 5 discusses more in-depth the role of timing and program intensity. Section 6 concludes and discusses future research.

2. Setting and Intervention

2.1 Setting

Despite considerable public spending on primary education, physical and human resources remain scarce across Indian primary schools (World Bank, 2013). The low number of teachers, typically lower than the number of grades in the schools, severely limits their ability to teach at the appropriate level of each child. Moreover, teachers are often more concerned about completing the demanding syllabus rather than about making sure every child reaches the minimum standard for their grade, given that in any case children are automatically promoted to the next grade (Banerji and Chavan, 2016).[4] When it comes to after-school activities, the key challenge is represented by the lack of opportunities for children to train and practice what they might have seen in class and the limited engagement of community and family members in the children's learning process (Banerji et al. 2017). These challenges are evident in Assam, which provides the setting for this study. Assam is a mountainous state, in the north-east of India and home to approximately 31 million people, distributed across 33 administrative districts. Data from our baseline, which collects information from 200 villages within Nagaon district (Figure 1), indicates that the five primary school grades on average share just 4 teachers and 2 classrooms. Moreover, only 74% of the children report studying after school and virtually no caregiver reports being involved in any after-school community-based studying and learning activity.

In Assam 97.7% of children in the age group of 6-14 are enrolled in school (ASER, 2018). The large majority (71.7%) is enrolled in public schools, while 24.8% attend private schools, and 1.2% attend other educational institutes (such as madrassas). Yet, despite near universal primary school enrolment, learning levels remain very low across the state. According to the most recent Annual Status of Education Report (ASER, 2018), only half of children in grade 5 can read a grade

[4] Until 2010automatic promotion was in effect until grade 5; the Right to Education Act, implemented in 2010, has extended the automatic promotion policy through grade 8.

2-level text or solve simple two-digit subtraction problems. Nagaon is the district where the NGO Pratham planned to expand its activities in 2018 and represents the specific setting of this study. It is the largest district of Assam in terms of population, with close to 3 million inhabitants, and according to the most recent disaggregated statistics, it performs on par with the average district in the state in terms of schooling outcomes (ASER, 2016).

2.2 Intervention and Implementation

The program studied in this paper was implemented by *Pratham Education Foundation,* the largest NGO working in the Indian education sector. Pratham was founded in 1995 to provide education to children in the slums of Mumbai. It has grown rapidly over time both in scope and size, developing a set of simple but innovative programs to improve the quality of learning for children across India. It is today operating in 23 States and union territories and reaches millions of children every year. Pratham started operating in Assam in 2007 and this study takes advantage of the expansion of its activities that took place in 2018.

Pratham is the pioneer of a pedagogical approach, called Teaching at the Right Level (TaRL), based on the simple yet powerful idea of re-grouping children in school not according to their age and grade, but rather by their actual ability level. The delivery of this approach through targeted learning camps inside schools was first evaluated in 2014 in Utter Pradesh, where it was found to significantly improve children's learning outcomes (Banerjee et. Al., 2016). While keeping this model as the central component, Pratham has repeatedly revised and refined its approach to adapt it to different contexts and to adjust it to ever changing circumstances. One persisting challenge Pratham encountered in its operation was the lack of opportunities for children to study and reinforce their learning outside of schools. While families and communities typically recognize the important of studying and learning, they often have limited time at disposal and no framework in place to provide support. To address this, Pratham recently developed a new community-based Study Groups program to be combined with its traditional TaRL approach. Here we provide more details about these two components.

2.2.1 Learning Camps

The *Learning Camps* consist of intensive bursts of teaching-learning activity, with a focus on the foundational skills for reading and arithmetic, administered by Pratham staff with the support of local volunteers, during school hours, when regular teaching is temporarily suspended and children are re-arranged in groups based on their ability. The camps are confined to three periods of 10 days each (for a total of 30 days), spread over one teaching term (5 months).[5] At the beginning of the first learning camp session, Pratham tests all children in grades 1 through 5, to identify their level of reading and arithmetic. Children are then grouped according to their level

[5] Before the beginning of the school term, Pratham also runs a 2-day event to introduce its programs to the community and to identify volunteers.

and taught language and mathematics for about 1.5 hours each by Pratham staff and Pratham-trained local volunteers. Teaching-learning activities and materials are tailored for each group and designed by Pratham. At the end of each 10-days camp Pratham assesses the progress of each child, making sure that he/she gets then allocated to the correct group for the next round. The hypothesis is that by receiving education appropriate to their actual level, children will be able to learn more and more rapidly. Even though the *Learning Camps* only last for 30 days in total, by learning basic skills, children are expected to reach a different learning trajectory and to learn more over time. At the same time, if the skills acquired during the camps are not supported by regular training, there is a risk that children will quickly revert to previous low learning levels.

2.2.2 Study Groups

The community-based *Study Groups* are set up in the villages by Pratham and are then managed by the community. The number of groups varies depending on the availability of volunteers to manage them. Each group is made up of about seven primary school age children and is coordinated by one volunteer (typically a family or community member). The coordinators are mobilized and the groups formed during an initial 2-day introductory event that Pratham holds in the village before the beginning of the school term. Participation is voluntary and there are no financial rewards for the coordinators or the groups. Once the groups are set up, a Pratham team member conducts monthly visits to the villages to share learning material that aims to guide and support the activities of the groups.[6] Groups do not necessarily have to focus on the material and are invited to work also on other reading and studying activities (especially homework). Overall, the structure of the groups and its activities are meant to be very flexible. It is up to the coordinator to set the frequency and length of the meetings.[7] The aim of the program is to provide a simple framework that family and community members can use to support primary school age children to practice and reinforce their learning outside of school, while keeping their own level of time engagement relatively low. In this way the program wants to encourage support in the community for children's learning, foster a culture of group and out-of-school learning among children, and stimulate a desire to read and learn so to enhance and sustain the skills that children acquire in school. The hypothesis is that the community-based Study Groups lead to an increase in the amount of time children spend studying together after school, which, in turn, is expected to lead to better learning outcomes.

---

[6] The material consists in a set of colored papers that contain readings, games, and exercises for different levels of ability. The material is meant to be self-explanatory and does not require the supervision of the group coordinator. In practice, the material is delivered by Pratham to a contact person in the village, who then shares it with the other group coordinators.

[7] Although groups did not maintain any formal records of their activities, after the conclusion of the study we conducted a short qualitative survey with 40 coordinators across the study villages. The information that we collected indicate large heterogeneity across groups. Most of the groups met 2-3 times per week for about 2 hours, and children mostly focused on the material shared by Pratham. However, the frequency and length of the meetings varied greatly, with some groups meeting every day.

3. Research Design

3.1 Randomization

We took advantage of the planned expansion of Pratham's activities in Nagaon district to randomly select villages to be reached by the program. The sample consists of villages where Pratham had never worked before and that it deemed eligible to host its activities. For the assessment, Pratham combined administrative data from Sarva Shiksha Abhiyan (SSA) – the Government of India's program for universal primary education – with in-field visits and defined a preliminary list of 260 eligible villages.[8] We then randomly selected 200 villages from this list to be part of the study. Figure 1 shows the location of Nagaon district, as well as the spatial distribution of the study villages within the district.

[Figure 1 about here]

After baseline data collection, the 200 villages were randomly divided into four groups (50 villages per group): *Learning Camps & Study Groups; Learning Camps*; *Study Groups*; *Control*. A second randomization was then performed, taking advantage of the fact that Pratham could not run the Learning Camps component at the same time across all target villages, due to budget and logistical constraints. Out of the 100 villages assigned to receive the Learning Camp (either alone or in combination with the Study Groups), we randomly selected (stratifying by group) half of them to receive the program right after baseline data collection (phase 1), and the other half to receive it roughly 5 months later (phase 2). Since the endline survey took place at the same time across the whole sample, this step introduced random variation in the time since exposure to the Learning Camp component. Figure 2 illustrates the experimental design.

[Figure 2 about here]

3.2 Data

We collect data for our analysis from three different actors: head teachers, children, and their caregivers (households).

Within each one of the 200 study villages, we surveyed the head teacher of the local primary public school.[9] The survey included a mix of direct questions and observational data to be recorded by the enumerators, focusing on school facilities, teachers availability, student enrollment and attendance, inputs availability, management practices, pedagogical methods,

[8] Pratham's standard preliminary assessment considers official enrolment in the primary school of the village, accessibility of the village, and a qualitative assessment of the potential for mobilizing the community. Note that, given that each village hosted one and only one primary public school, we can refer to village or school interchangeably.

[9] As mentioned above, in the study area each village included one and only one primary public school.

and support from NGOs and government.

We then selected a representative sample of children enrolled in grades 1 to 4.[10] At baseline we targeted 8 randomly chosen students per grade, for a total of 32 students per school.[11] Student selection was based on school enrolment registries and not on actual attendance, so to avoid missing students that were absent from school on survey day. Selected students that were absent from school were located and surveyed at home. The child survey consisted of two parts: a test and a short survey. The test component was itself divided into two tests, which we labelled "Test A" and "ASER test". Test A was created by the research team and based on tests previously used in other studies in India (Muralidharan et al. 2018). The test was especially designed to target children in the lowest grades and contained basic math and language questions. The second test mirrored the standard ASER test - a nationwide test, divided in a math and a language component, that is conducted yearly by the ASER Center all over India for children aged 5 to 16. Both Test A and ASER were conducted individually with each child. Appendix B provides richer details about the tests. Unfortunately, by endline test A proved to be too easy for the students in the sample and we faced severe top-coding issues, especially in the language section. For this reason, in our main analysis we focus on the ASER test, where we have more variation, and we report the result on test A in Appendix A.[12] The short survey was administered after the tests and contained questions on study habits outside of school, time allotted to different activities in a day, perceptions about learning, aspirations for the future, and psychological wellbeing (this last component was only included in the endline survey).

Finally, we also surveyed a representative sample of caregivers of the children included in our study. At baseline we targeted 24 children (6 per grade) and surveyed their primary caregivers at home.[13] The respondent was identified as the person who dedicates the most time to look after the needs of the child. In the large majority of cases this was the mother (85%), followed by father (5%), and grandmother (3%). The survey collected information on household characteristic, educational expenditures, non-financial educational inputs (e.g. helping with homework), time allotted to different activities in a day by both the caregiver and the child, and aspirations for the

[10] The choice of excluding students enrolled in grade 5 at baseline was made to ensure proper program exposure, as children enrolled in grade 5 at baseline would have completed primary school within the following few months and likely moved to a different village to continue secondary studies, as it is common practice in this setting.

[11] If enrolment in a given grade was higher than 8, supervisors randomly selected the children to be surveyed using a randomization table created by the research team. If enrolment in a given grade was lower than 8, additional students were surveyed from other grades in the same school, following a pre-determined selection procedure. In the end, actual enrolment figures were lower than official figures included in the SSA administrative and therefore, the final child sample surveyed were on average 29 children per school, or about 7 children per grade per school (see table A.1 in appendix for the breakdown by grade).

[12] Results are very consistent for the mathematics section, while they are instead typically non-significant for the language section, where the top-coding issue was more severe.

[13] Because some schools were smaller than expected, we managed to complete an average of 23 primary caregivers surveys in each village (see table A.1 in appendix for the breakdown by grade of the corresponding child).

child's education.

Figure 3 presents the timeline of the project, with implementation-related activities listed below the line, and research-related activities above the line. The baseline survey was conducted between May and August 2018. Implementation of the programs was done by Pratham and started immediately after baseline data collection. As discussed above, the implementation of Learning Camps was rolled out in two phases: half of the schools assigned to receive the camps hosted them between August and December of 2018 (phase I), while the other half hosted them between January and May of 2019 (phase II). The community Study Groups started in all 100 target villages in August 2018 and continued until the end of the study. A short compliance survey was conducted between May and June 2019. The endline survey was conducted between November 2019 and January 2020, roughly 16 months after the start of program implementation. The trial was registered in the AEA RCT registry (number 0002817) and received ethical approval from IFMR Human Subjects Committee (IRB00007107) and TCD Ethic Review Board (05062018).

[Figure 3 about here]

3.3 Summary Statistics and Validation

Table 1 shows that the randomization led to the creation comparable groups at baseline: observable characteristics of schools, children, and caregivers are well balanced across treatment arms, as are normalized test scores in each subject.[14] Table 1 also provides summary statistics on the study population.

[Table 1 about here]

The primary schools in our study are small, with an average of about 2 classrooms and 4 teaching staff members (including head master, teachers, and para-teachers) to take care of the five primary school grades. Total enrollment at baseline was close to 53 students per school, equally divided between girls and boys. On average children in our sample were 7.5 years at baseline, with a perfect split between boys and girls (Panel B). About 70% of the children were present in school on survey day. Children overall reported liking school: their average rating was more than 4.5 on a 5-point Likert scale.[15] On average, only about 75% of children reported ever studying after school. Looking at ASER test scores (our primary outcome of interest), the average student in our sample at baseline scored between level 1 (beginner) and level 2 (corresponding to 1-digit number recognition in math or letter recognition in language) in both test components.

[14] Table 1 only includes a subset of the variables collected at baseline. When we consider the entire set of variables at disposal and perform baseline checks across groups, we have 213 total comparisons. Out of these, we observe 25 instances (11.6%) in which the difference is significant at 10% level (p-value<0.1), 14 instances (4.5%) in which it is significant at 5% level (p-value<0.05), and only 1 instance (0.5%) in which it is significant at 1% level (p-value<0.01).

[15] Informed by extensive piloting, the question was administered using visual aids: the enumerators showed to the children five stylized faces, that ranged from very sad (1) to very happy (5) and asked the children to indicate their answer on that scale.

This average obviously hides large heterogeneity by grade (Figures A.1 and A.2 in appendix). While it is clear that there is progression across grades, the overall learning levels remain well below the expected standards. In math, for instance, the share of students who could recognize 2-digit numbers increases from 11% in grade 1 to 65% in grade 4. Still, only 3% of students in grade 4 were able to complete divisions (which is the expected standard for a grade 3 student) and 35% were not even able to complete subtractions (which is the expected standard for a grade 2 student). The picture is similar when looking at language: only 8% of grade 2 students were able to read a story – which is the expected standard for that grade – and the share increased to just 21% and 34% in grades 3 and 4, respectively. Finally, on average the households included in our sample had slightly more than 5 members, and out of these, usually between 1 and 2 were enrolled in school. Only half of the primary caregivers defined themselves as literate, while they held high hopes for the education their children: 45% of them hoped their child will reach university. About 20% of them paid extra school tuition to support their children's learning. After explaining how the ASER test works, we asked the caregivers to assess their children's level and overall, only about a third guessed the correct level (33% in language and 38% in math), while the large majority overestimated the ability of their child (53% in language and 55% in math).

Attrition at endline was low and we were able to track back 93% of the baseline sample. The attrition rate was similar across all study arms and the characteristics of children lost at endline did not differ across study arms (table A.2 in appendix).[16] These checks alleviate concerns related to potential differential attrition and make us confident in attributing any difference that we might observe in the outcomes at endline to the program.

3.4 Empirical Strategy

Our main estimating equation for child-level outcomes takes the form:

$$\mathrm{Y}_{i,v} = \alpha_1 LC\&SG_v + \alpha_2 SG_v + \alpha_3 LC_v + \Omega X_{i,v} + \varepsilon_{i,v} \qquad (1)$$

where $\mathrm{Y}_{i,v}$ is the outcome of child i, living in village v. $LC\&SG, SG$, and SG, are indicators taking on the value of one if the village was assigned to the full program (which included both Learning Camps and Study Groups), to the Study Groups component alone, or to the Learning Camps component alone, respectively. In order to increase precision of the estimate, we also include in the regression a vector $\mathrm{X}_{i,v}$ of control variables defined at the individual level: age, gender, an indicator for each grade, and the baseline value of the outcome variable $\mathrm{Y}_{i,v}$. We cluster standard errors at the village level, accounting for the fact that the intervention varies at that level. Coefficient α_1 shows the

[16] At endline survey we were not given permission to conduct one school survey, as the head teacher of that school (in the control group) refused to answer the endline questionnaire. Caregivers' attrition mirrors children's attrition (93%): conditional on successfully tracking back the child, at endline we managed to conduct the caregiver survey in every household that was included in the baseline caregiver sample, although in 164 cases (3.8%) the baseline respondent was not available and was replaced with another caregiver from the same household.

impact of the full program on the outcome Υ, while α_2 and α_3 capture the impact of the two individual components – Learning Camps and Study Groups, respectively – whenever they are implemented alone. By comparing these three coefficients to each other, we can also test whether the effectiveness of the program differs across the three treatment groups. The analysis at the school and household level follows a similar specification as (1), with variables defined at the corresponding level.

To form judgment about the impact of the intervention on a family of n related outcomes, and address potential multiple hypothesis testing concerns, we follow Kling et al. (2004) and estimate a seemingly unrelated regression system

$$\Upsilon = [\mathrm{I}_n \otimes LC\&SG]\alpha_1 + [\mathrm{I}_n \otimes SG]\alpha_2 + [\mathrm{I}_n \otimes LC]\alpha_3 + \mu \qquad (2)$$

where Υ is a vector of n related outcomes, I_n is an n by n identity matrix, and $LC\&SG, SG$, and SG, are vectors of assignment to treatment groups indicators. For each coefficient α_i we derive an average standardized treatment effect (ASTE)

$$\widetilde{\alpha_\iota} = 1/n \sum_{n=1}^{N} \widehat{\alpha}_{i_n}/\widehat{\sigma}_n \qquad (3)$$

where $\widehat{\alpha}_{i_n}$ is the point estimate on one treatment indicator in the nth outcome regression and $\hat{\sigma}_n$ is the standard deviation of the control group for outcome n (see Duflo et al. 2007).

4. Results

4.1 Program Implementation

As a first step, we use information contained in the endline child and caregiver surveys to check self-reported awareness of and exposure to the intervention programs. While during the implementation phase we used Pratham's rich administrative records to confirm that the study design was fully respected, information contained in the child and caregiver surveys provide us with information related to program implementation from the perspective of the ultimate beneficiaries. Panel A of Table 2 reports the results based on children's responses. Although few students reported having heard about Pratham (on average 18% across the three treatment groups, column 1), the comparison across groups is in line with what we expected based on the study design: children in villages assigned to the Learning Camps program were much more likely to report participating in camp-related learning activities in school (columns 2-4), while children in villages assigned to the Study Groups activities were significantly more likely to report that their village hosted study groups (column 5) and that they themselves took part in them (column 6). When we combine the different measures capturing program-related activities, we clearly see that children in treatment villages report significantly higher interactions with Pratham (column 7) and exposure to the specific programs targeting their villages (columns 8 and 9), compared to children in control villages.

Panel B of Table 2 relies instead on information collected from caregivers and the results appear overall in line with those based on children's responses. Only a small portion of the caregivers in treatment villages reported having heard about Pratham (column 1) or having interacted with it (column 2), although about a third recognizes a sample of teaching and learning material that Pratham regularly uses for its activities (column 3). Columns 4 to 9 show that caregivers in the treatment villages are significantly more likely to report activities related to Pratham's programs taking place in their village or school. When comparing outcomes across treatment arms, the answers concerning the presence of study groups well reflect the study design (column 7), while it is less clear for answers related to the learning camps (columns 4 – 6 and 9), which is likely a consequence of the fact that caregivers are less aware of what happens within the classrooms.

[Table 2 about here]

Taken together, results in Table 2 confirm that the study design led to a significant change in the pedagogy in school and in the presence of study groups across villages assigned to the treatment arms. At the same time, only a fraction of the target beneficiaries reports direct participation in the programs. This can be partly justified by the fact that that children are unlikely to be fully aware of the distinction across various teaching and learning activities and might not correctly remember and/or report their participation. Similarly, caregivers might not be fully aware of the activities going on in the school and community. These self-reported measures should therefore be considered lower-bound of direct participation in the programs. Nevertheless, these figures clearly suggest that not every child in treatment villages got directly exposed to the programs. Some children might simply miss school when Learning Camps are running or might not attend the study groups organized in the village. Results in the following subsections should therefore be interpreted as measuring the impact of having the programs implemented in the child own school or village, and *not* the impact of directly participating in them. We will discuss more in detail the role of direct exposure in section 5.

4.2 Learning Outcomes

We report the average treatment effects on learning outcomes – our primary outcome of interest – in Table 3. The table shows that the full education program, which combines Learning Camps and Study Groups, led to a significant increase in both mathematics and language test scores.[17] The estimates imply that children in villages assigned to the standard Pratham program experienced on average 0.09-0.12 standard deviation increase in mathematics and language, compared to children

[17] As discussed above, given the issues we encountered with top-coding in Test A, we focus our main analysis on ASER tests. Results based on test A are very consistent for the mathematic section, while typically not-significant for the language section (which suffered severely of top-coding). Results are reported in Table A.3 in Appendix A.

in the control group.[18] These effects translate in a significant increase in the share of children that are able to achieve minimum standards (i.e. able to perform at grade-2 level) in the two subjects: a 20% (or 6pp) increase in math and a 13% (or 4pp) increase in language. [19]

[Table 3 about here]

When we study at the impact of the two program components individually, instead, we do not find any impact, irrespectively of the subject considered. Not only the results are not statistically significant, but the estimated coefficients are close to zero. The p-values reported at the bottom of the table confirm that implementing the full program led to a significantly higher impact on test scores compared to implementing either of the two components alone (although when we look at the coarser measure of students achieving minimum standards, we lose some power and the differences are mostly not significant at conventional levels). Figure A.3 in Appendix visually illustrate the change in learning levels experienced across the different study arms, while Figure A.4 and A.5 report the distribution of the different learning levels by grade and study arm.

4.3 Schooling and Studying Practices

In order to better understand the impact of the programs, we look at a richer set of children-related outcomes that we collected through our survey. Results are reported in Table 4. First, we see that the programs had no impact on school attendance, neither when measured by enumerators through presence in class on survey day (column 1), nor when based on the caregivers' information (column 2).[20] This indicates that the impact on test scores discussed above are not affected by changes in exposure to regular teaching.

Next we consider children's study habits. In line with what we observed in Table 2, we find that the presence of Study Groups in the village led children to study significantly more after school (column 3) and, in particular, led them to study more in groups with their friends (column 4): this share increased from virtually zero to 1.2% and 2.7% in the study group and full program treatment arms, respectively. These shares are still allegedly very low, even when allowing for under-reporting, and suggest that only a small fraction of students directly participated in the groups. Unfortunately, we cannot rigorously check participation, as the groups were managed and organized by the local community that did not keep any formal records. Nevertheless, in section 5 we will provide a suggestive analysis of the role of exposure to the program, by exploiting Pratham's own records on the number of groups that were created in each village.

---

[18] The structure of the ASER test in levels is not ideal for expressing results in terms of standard deviations, but we nevertheless provide the estimates in SD to facilitate comparability with other settings. At endline the observed SD in math and language scores in the Control group are equal to 0.948 and 1.389, respectively.

[19] As mentioned above, grade-2 level means being able to solve a subtraction problem (math) or to read a short story (language).

[20] Similar results are obtained when considering enrollment and attendance from the school registries.

Finally, we study the impact of the programs on children's schooling ambitions, school and learning perceptions, and psychological wellbeing, by grouping together families of related outcomes and computing average standardized treatment effect (ASTE) as described above.[21] The results reported in columns 5 to 9 show no clear pattern, indicating that the intervention had no impact on these dimensions.

[Table 4 about here]

4.4 Caregivers' Routines and Behaviors

Caregivers were not a direct target of the programs, which, as explained above, were purposely designed to minimize the time commitment required from any family or community member. With the rich data at our disposal, we can check whether there was any indirect impact on their routines and on their engagement in the education of their children.

Any potential impact on caregivers is a priori ambiguous, as they could consider new teaching and learning programs as substitutes or complements of their own involvement in the education of their children, thus leading to decrease or increase their engagement. One might also expect that any impact on caregivers' behavior and schedule, if at all present, is more likely to come from the community Study Groups, which take place after school, rather than with the Learning Camps, that instead take place in school during regular school hours. Although also in this case the impact is ambiguous, as on the one hand, the Study Groups might free up some time for the caregivers, thanks to the fact that children can now study with their peers. On the other hand, some caregivers might see the community putting more weight on after-school learning and hence increase the time they dedicate to support their children also at home. Results reported in Table 5 shows that caregivers in villages that organized the Study Groups were significantly more likely to report being involved in after-school learning groups, with the share moving from virtually zero to about 3 percent in the villages that hosted the groups (column 1). However, we do not observe any significant change in the average amount of time caregivers spends studying, reading, or playing with their child on a regular day (column 2) or in the hours per week that children spend doing homework with a

[21] To capture ambitions, we asked until which grade the child would like to study. To measure school and learning perceptions we asked three sets of questions: 1) we directly asked the importance they attach to education; 2) we asked how much they like school, reading, and mathematics; 3) we asked a set of 8 questions about perception of schooling and learning. To capture psychological wellbeing we relied on a tool recently developed by a team of researchers based in the school of education at University College Dublin. The tool is called Child and Adolescent Social and Personal Assessment of Wellbeing (CAPSAW) and has been extensively piloted and validated over the past 3 years in different contexts (Ireland and Sierra Leone), on children of age 5 to 16. The full tool considers 4 domains and contains 8 questions for each one of them. We focused on the two domains of interest for our study: personal well-being and teacher-related well-being. The individual questions capture how much children agree with statements such as "I feel safe with my teachers", "Teachers help me if I have a problem", "People generally care about me" and "I am helpful to other people". For each family of related outcomes, we estimated the seemingly unrelated regression system (2) to derive the average standardized treatment effects (ASTE), which is reported in the table.

household member (column 3).[22] Similarly, there is no change in the household expenditure on the education of the child (column 4). In other words, it seems that other than leading some caregivers to get involved in the study groups, none of the interventions led to any clear re-adjustment in the household time and financial investment in children's education at home.

By intervening on what happens both inside and outside of school, however, the programs might have impacted the caregiver's interactions with the school. In column 5 we report the average standardized treatment (ASTE) effect obtained, as described in section 3.4, from a seemingly unrelated regression system, which includes five different variables related to parents-school interactions.[23] The results indicate that all programs significantly increased caregivers' interactions with the school.[24] This result is quite remarkable, as it is notoriously difficult to increase parental engagement with the schools, even with high-intensity programs specifically targeting parents (e.g. Di Maro et al, 2020).[25]

Finally, we also study whether the programs impacted the caregivers' knowledge about their children's ability in math and language, using the ASER scale. We find that at endline less than half of the caregivers correctly estimated the learning level of their child (only 45% of caregivers correctly estimate their children's language level and 41% correctly estimate the math level), and the overwhelming majority of the mistakes comes from an overestimation of their child (75% for language and 90% for math). However, we do not find any evidence that the programs corrected caregivers' misaligned knowledge of their children's learning levels.

[Table 5 about here]

4.5 Schools' Activities and Investments

Finally, we turn to look at the impact of the programs on the resources and practices of the primary public schools located in the study villages. Also in this case the impact is a priori ambiguous and depends on whether schools (and authorities monitoring the schools) consider Pratham's intervention as a complement or substitute of their standard resources and activities.

---

[22] In order to record time use, we adopted a well-established approach, which we also extensively piloted in the field: we gave respondents a set of beans and asked them to place them on a number of images depicting different activities, explaining that more beans indicated more time spent on that activity.

[23] The five variables capture whether: 1) the caregiver attended parent-teacher meeting, 2) the caregiver received a call from a teacher, 3) the caregiver received information from the school, 4) the caregiver discussed the child's education with the school; 5) the caregiver checked the child's marks at school.

[24] When we disaggregate the index, we find that across all treatment groups caregivers are more likely to report attending a parent-teacher meeting (statistically significant for the SG and LC treatment arms), to have received a call from a teacher (statistically significant for LC & SG and LC), and to have received information from the school (statistically significant for LC). We do not find instead any clear evidence that caregivers in any treatment groups were more likely to discuss education with the school, or to check their children's marks.

[25] The fact that more parents-schools interactions do not translate in higher test scores in the SG and LC groups is in line with recent evidence from Barrera-Osorio et al (2020) in the context of Mexico.

We start by looking at the schools' human and physical investments. Results are reported in Table 8. Columns 1 and 2 show that the programs did not have any impact neither on the number of teaching staff members appointed to the school, nor on their attendance rate. Columns 3 to 6 look instead at the impact on physical investments in the school during the study period. The variables focus on investments in constructions (column 3), reparations (column 4), teaching and learning material (column 5), and seating mats for the students (column 6). Column 7 provides the ASTE across these four outcomes and indicate a reduction in overall physical investments across all treatment groups, with the drop in the Learning Camps group large and significant at 5%.

Finally, we consider schools' management practices, by looking at the frequency of the meetings of the school management committee (column 8), the parent-teacher association (column 9), and institutional meetings with the Ministry of Education (column 10). Column 11 provide the ASTE estimate across the three measures, and clearly indicate that schools that hosted the Learning Camps experienced a lower frequency of meetings with stakeholders.

These results indicate that Pratham's activity in the schools to implement the pedagogical intervention induced a drop in physical investments in the schools as well as a decrease in their interactions with stakeholders. To the extent that these represented relevant inputs to support the children's learning process in school, such substitution might have undermined the effectiveness of the program over the study period.

[Table 6 about here]

5. Discussion

5.1 Evolution Over Time

In its review of the literature, McEwan (2015) finds that only 1 in 10 education studies was assessed more than a month after the interventions ended, thus making it impossible to understand how learning outcomes evolve post-intervention. Our study design allows us to shed some light on this dimension, at least for what concerns the Learning Camps component. As explained above, the camps were implemented in two different phases due to cost and logistic reasons, and assignment to phase I or phase II was randomized by the research team. This generated exogenous variation in the time lag between the conclusion of the Learning Camps and our endline survey, which ranged between 6 and 12 months. We therefore run a regression similar to regression (1), where we split the LC&SG and LC indicators in two variables each, to distinguish between the two phases. Results reported in Table 7 indicate that being exposed to the Learning Camps earlier on led to higher test score by endline. The coefficient of the full intervention indicator is twice as large for students that received the program in phase I as opposed to those who received it in phase II, although the difference is not statistically significant at conventional levels. Figure A.6 in Appendix visually illustrates the evolution in learning across the different study groups.

This result is consistent with the hypothesis that, by focusing on foundational skills, Learning Camps manage to put children on a different learning trajectory, widening the gap with the control group over time. Interestingly, this pattern only emerges for the full intervention, where the pedagogical component is complemented by the community study groups, indicating that the groups provided the children with an important opportunity to strengthen their learnings and take advantage of the new learning trajectory.

[Table 7 about here]

5.2 Program Exposure

Our analysis so far focused on "intention-to-treat" effects, as it did not take into account whether a child actually attended the Learning Camps or the community Study Groups organized in their schools and villages. One relevant question is what is the impact of *direct* exposure to the programs. Although we cannot perfectly identify such "treatment-on-the-treated" effect within our setting, in this section we provide some suggestive evidence, exploiting two sets of administrative data collected by Pratham as part of its standard practice. First, concerning the Learning Camps, we make use of the complete records of students that attended the camps. We define our direct exposure measure as the share of Learning Camps days (out of 30) that the student attended (average among children in the Learning Camps study arms is 68%).[26] Second, concerning the community-based Study Groups, we used information on the total number of groups active in each villages. Ideally we would like to know exactly who attends each group, but the communities and coordinator did not keep any formal records. We therefore use the number of groups, weighted by the number of students enrolled in the primary school at baseline, to create a proxy for exposure to the Study Groups (the average is 0.1, meaning one group per every 10 children enrolled in primary school).

Columns 1 and 4 in Panel A of Table 8 report the results on math and language scores when we replace assignment to the learning camps study arms with the measure of attendance to the camps described above. Compared to findings reported in Table 3, these estimates indicate significantly larger gains for children that were fully exposed to the camps, both when camps were implemented in combination with the study groups, but also when they were implemented alone. Columns 2 and

[26] For generating this measure we proceed by step, by first matching the children in our sample with Pratham's records and then computing the learning camps attendance measure for each child. Out of the total 2,760 children in our Learning Camps sample, 93% were successfully matched with Pratham's administrative records. Although we cannot fully rule out some omissions and entry errors, we feel confident that unmatched children are not simply caused by minor spelling inconsistencies. We combined an automated fuzzy matching program with manual checks and refinements to ensure the accuracy. The matching took into account information on school, grade, age, gender, first and second names. The share of children that could not be matched (i.e. never attended the camps) was similar across LC&SG and LC groups (p-value=0.717).

4 of Panel A report instead the results when replacing assignment to the study groups study arms with the measure of density of Study Groups operating in the village described above. In this case results are in line with our previous finding that Study Groups can induce significant improvements in learning outcomes when combined with the Learning Camps, but not when implemented alone, and show that larger gains are realized in villages with a higher density of Study Groups. Finally, columns 3 and 6 confirm these results, when both the new measures of program exposure to the Learning Camps and Study Groups are considered in the same regression.

An obvious concern with the analysis in Panel A, is that while the allocation of the programs across different villages was random, program exposure was not and is likely driven by a number of characteristics associated with learning outcomes. It is indeed possible that students that attended more learning camp days were simply more motivated students that would have performed better anyway. Similarly, villages with higher density of Study Groups might have been more supportive to children's learning even in the absence of the program. Both these factors would suggest that estimates in panel A are overestimating the impact of exposure to the programs. In an attempt to address this endogeneity, we resort to an instrumental variable approach, using random assignment to the different study arms as our instruments. This analysis should be viewed suggestive, as it rests on the assumption that the only way in which the programs affected the learning outcomes is through direct participation in the Learning Camps (for the Learning Camps component) and/or through the number of Study Groups created in the village (for the Study Groups component).[27] With this caveat in mind, panel B of Table 8 reinforces our earlier results, showing that higher exposure to the program in the full program group (which is the only group where our "reduced form" results in table 3 showed a significant effect) leads to significantly higher test scores. Results in columns 3 and 6 indicates that by increasing learning camps attendance from 10 to the full 30 days a child located in a village that hosted the full program, with an average number of Study Groups per child (equal to 0.1), could boost her math and reading scores by 0.1 points (0.11 SD) and 0.12 points (0.08 SD) respectively.

[Table 8 about here]

5.3 Heterogeneity Analysis

We also study whether the interventions had differential impact depending on the child gender, grade, starting learning level, and baseline school quality. In none of these cases we observe any clear pattern, indicating that the program had similar effects irrespectively of these characteristics.[28]

[27] This assumption would be violated for instance if Pratham's presence in the villages affected children's learning outcomes also by changing the overall attitude towards teaching and learning in the school and community, or if children not directly exposed to the programs were still affected through their interactions with children that participated in them.

[28] Results are reported in Table A.4 in Appendix.

5.4 Cost Effectiveness

In this section we provide a cost effectiveness analysis for full program, for which we observed a significant impact on student's learning outcomes. This analysis has two objectives. First, it discusses the cost implications of implementing the full intervention as opposed to just one of its components. Second, it provides a metric that allows comparisons with other interventions assessed in the literature.

In order to perform an accurate cost analysis, we worked with Pratham to define three alternative scenarios, mirroring the three intervention arms, and defined the costs associated to each one of them. In practice, this meant detailing the yearly costs that Pratham would incur if it was to serve 50 villages with one of the three interventions studied here – the full program, Learning Camps only, and Study Groups only. Given that the study period lasted 16 months, we make the simplifying assumption that costs linearly increase over time and multiply the yearly costs by 1.33 to obtain the costs relative to the study period. Table 9 reports the summary of the main budget items across the three scenarios (in INR).

There are three clear indications emerging from the table. First, the Learning Camps component is much more expensive than the Study Groups component. Second, the personnel cost is by far the largest item in each scenario, ranging between 67% and 75% of total costs, followed by the creation of the teaching and learning material (TLM) used in the programs, which ranges between 11% and 17% of total costs. Third, there are very strong synergies that can be exploited by combining the two components. The most important one comes from the fact that the same people that run the Learning Camps in the schools can easily work with the community to set up the Study Groups program. Overall, the average cost is 1,090 INR per student for the Learning Camps only program, 681 INR per student for the community Study Groups only program, and 1,176 INR per student for the full program that combines the two.

In order to provide a single summary measure of cost effectiveness and make this analysis easily comparable to other studies in the literature, we estimate how much it would cost on average to raise learning outcomes by 0.1 SD.[29] Estimates are based on results from Table 3. The cost of raising learning outcomes by 0.1 SD with the full program ranges between 980 INR (13.7 US$) and 1,306 INR (18.3 US$) per student. Stated otherwise, under the (greatly) simplifying assumption of linear returns, an investment of 100 US$ per student would lead to an increase in learning outcomes of about 0.55 SD in reading and 0.73 SD in mathematics. To put these numbers in perspective, we consider a set of 27 education programs aimed at improving learning outcomes that were evaluated using randomized controlled trials and for which J-PAL estimated cost effectiveness (Figure A.7 in Appendix). Out of the 27 programs, 15 found a significant impact, ranging between 0.13 SD and 0.59

[29] As explained above, the structure of the ASER test is not ideal for expressing results in terms of standard deviations, but we nevertheless provide the estimates in SD to facilitate comparability with other settings.

SD (mean 0.24 SD). Cost effectiveness, expressed in terms of additional gains per 100 US$ invested, varies greatly, ranging between 0.06 (from a conditional cash transfer in Malawi) and 118 (from an information intervention in Madagascar).[30]

[Table 9 about here]

6. Conclusion

In this paper we studied a primary education program designed to improve children's learning process both in-school and out-of-school. The results show that the joint implementation of the two components is key to attain significant improvements in learning outcomes. Learning is a multi-dimensional process and our findings indicate that educational programs that intervene on different dimensions at the same time have higher chances of taking advantage of the synergies and of leading to significant learning gains. In particular, our findings shed some new light on the importance of studying practice out-of-school, a dimension that has received relatively less attention in the literature, showing that a relatively simple and cheap intervention can greatly enhance the effectiveness of a more costly school-based intervention. This can help the design of programs and policies by government agencies and organizations working in the education sector. Pratham already made important steps in this direction and has adopted the full program studied in this paper as its new flagship educational model.

There are two important additional considerations emerging from our analysis. The first one relates to the input substitution that we observed in the schools where Pratham implemented the pedagogical component. Such input substitution likely undermined the effectiveness of the program and is a challenge that many education program are likely to face: in a context with scarce resources, the provision of direct support from an external entity might induce local actors, such as schools and local authorities, to substitute away some of their costly inputs. With the data at our disposal we cannot not fully investigate the mechanisms behind this finding, which deserves further attention in future work, but such behavior appears consistent with what Das et al (2013) have observed among households, who substitute away some of their own costly inputs into the education of their children, whenever the local school receives extraordinary grants injections.

Secondly, our analysis highlight the importance of timing. Our results suggest that the pedagogical intervention that took place in school succeeded in placing students on a different learning trajectory, with the gains becoming more evident over time, as students keep accumulating new knowledge. This suggests that very short-term evaluations of education programs, which, as

[30] Our estimates also appear within the range of what other studies have found concerning successful pedagogical intervention in South America. Across a range of ten studies implemented between 2009 and 2015 in Argentina, Belize, Paraguay, and Peru, the average cost of a 0.10 standard deviation increase in math test scores was 14.53 US$ per student, ranging between 6.90 US$ in Argentina (2009) and 22.48 US$ in Paraguay (2013).

described by McEwan (2015) in his review, represent the norm in the literature, are likely to provide only incomplete assessments. One obvious follow-up question is how the learning gaps might evolve over even longer time horizons. And a related question is what happens to the learning gains when students are unable to attend school for a prolonged period of time, such as during the prolonged lockdown associated to the COVID-19 pandemic. Whether and how these education programs are able to cushion the negative consequences of the lockdown are open questions left for future research.

References

Aoki, Aya. 2005. "Assessing learning achievements and development impact: Ghana's national functional literacy program." Australian Journal of Adult Learning, 45(1): 63(81).

Altinok, N., Angrist, N., & Patrinos, H. A. 2018. Global data set on education quality (1965–2015). The World Bank Policy Research Working Paper 8592.

ASER. 2016. Annual Status of Education Report (Rural) 2016. http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%202016/aser_2016.pdf Pratham Organization.

ASER. 2018. Annual Status of Education Report (Rural) 2018. http://img.asercentre.org/docs/ASER%202018/Release%20Material/aserreport2018.pdf Pratham Organization.

Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A., & K. Geven. 2020. Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes: A Set of Global Estimates. *Policy Research Working Paper* 9284. World Bank, Washington, DC.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. & Walton, M. 2016. *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India*. NBER working paper series w22746.

Banerjee, A., & Duflo, E. 2012. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

Banerji, R., & Chavan, M. 2016. Improving literacy and math instruction at scale in India's primary schools: The case of Pratham's Read India program. *Journal of Educational Change*, *17*(4), 453-475.

Banerji, R., Berry, J., & Shotland, M. 2017. The impact of maternal literacy and participation programs: Evidence from a randomized evaluation in india. *American Economic Journal: Applied Economics*, *9*(4), 303-37.

Barrera-Osorio, F., Gertler, P., Nakajima, N. & Patrinos, H. 2020. Promoting Parental Involvement in Schools: Evidence From Two Randomized Experiments. NBER working paper series w28040.

Bhula, R., Mahoney, M. and K. Murphy. 2013. Conducting cost-effectiveness analysis (CEA), J-PAL publication, available at https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and V. Sundararaman. 2013. School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5 (2): 29-57.

Duflo, E., Glennerster, R. and Kremer, M. 2007. Using Randomization in Development Economics Research: A Toolkit. In *Handbook of Development Economics*, Vol. 4, edited by T. Paul Schultz and John A. Strauss, 3895–3962. Amsterdam: North-Holland.

Ganimian, A. J., & Murnane, R. J. 2016. Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, *86*(3), 719-755.

Glewwe, P., Kremer, M., & Moulin, S. 2009. Many children left behind? Textbooks and test scores in Kenya. American Economic Journal: Applied Economics, 1(1), 112–35.

Glewwe, P., & Muralidharan, K. 2016. Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education* (Vol. 5, pp. 653-743). Elsevier.

Kling, J.R., Liebman, J. B., Katz, L. F., and Sanbonmatsu, L. 2004. Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment. *Princeton University Industrial Relations Section Working Paper* 481.

Lakshminarayana, R., Eble, A., Bhakta, P., Frost, C., Boone, P., Elbourne, D., & Mann, V. (2013). The Support to Rural India's Public Education System (STRIPES) trial: A cluster randomised controlled trial of supplementary teaching, learning material and material support. *PloS one*, *8*(7), e65775.

Lai, F., Luo, R., Zhang, L., Huang, X., & Rozelle, S. 2015. Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. *Economics of Education Review*, *47*, 34-48.

Linden, L. L. 2008. *Complement or substitute?: The effect of technology on student achievement in India*. Working Paper, Columbia University: InfoDev.

McEwan, P. J. 2015. Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353-394.

Muralidharan, K. 2017. Field experiments in education in developing countries. In *Handbook of economic field experiments* (Vol. 2, pp. 323-385). North-Holland.

Muralidharan, K., Singh, A., & Ganimian, A. J. 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, *109*(4), 1426-60.

Romero, M., Chen, L., & Magari, N. 2019. Cross-Age Tutoring: Experimental Evidence from Kenya. Working Paper.

United Nations, General Assembly. 2015. *Transforming our world: the 2030 Agenda for Sustainable Development*, A/RES/70/1. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

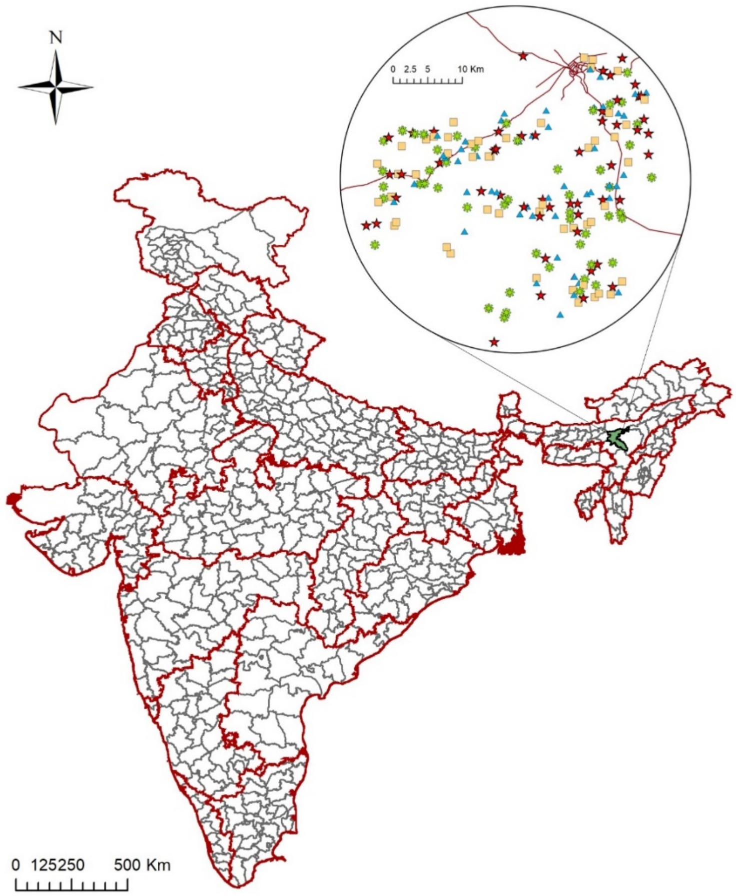
World Bank, Education Statistics (EdStats). 2013. Government expenditure on education as % of GDP (%). [Data file.] http://datatopics.worldbank.org/education/wDashboard/dqexpenditures

World Bank. 2017. *World Development Report 2018*. World Bank Publications.

World Bank. 2018. *School enrollment, primary (% net) - Low & middle income* [Data file.] https://data.worldbank.org/indicator/SE.PRM.NENR?locations=XO

Figures and Tables

Figure 1. Study Locations



Note: The study was set in Nagaon district, within the State of Assam. We drew a sample of 200 villages from the list of villages that Pratham deemed eligible to receive the program. The villages were then randomly assigned to one of the four study arms (50 villages per arm). The map above illustrates the location of Nagaon district within India, as well as an enlarged view on the location of the 200 study villages. The four different symbols indicate assignment to the different study arms.

Figure 2. Experimental Design

<table>
<tr><td colspan="3"></td><td colspan="2">Study Groups</td></tr>
<tr><td colspan="3"></td><td>No</td><td>Yes</td></tr>
<tr><td rowspan="3">Learning Camps</td><td colspan="2">No</td><td>C
(50 villages)</td><td>T1
(50 villages)</td></tr>
<tr><td rowspan="2">Yes</td><td>Phase I</td><td>T2a (25 villages)</td><td>T3a (25 villages)</td></tr>
<tr><td>Phase II</td><td>T2b (25 villages)</td><td>T3b (25 villages)</td></tr>
</table>

Note: The 200 villages were randomly assigned to one of the four main study arms as shown in the above table. Villages assigned to receive the Learning Camps (T2 and T3) were further randomized in two groups to determined when they would receive the program (see Figure 3 for timeline).

Figure 3. Timeline

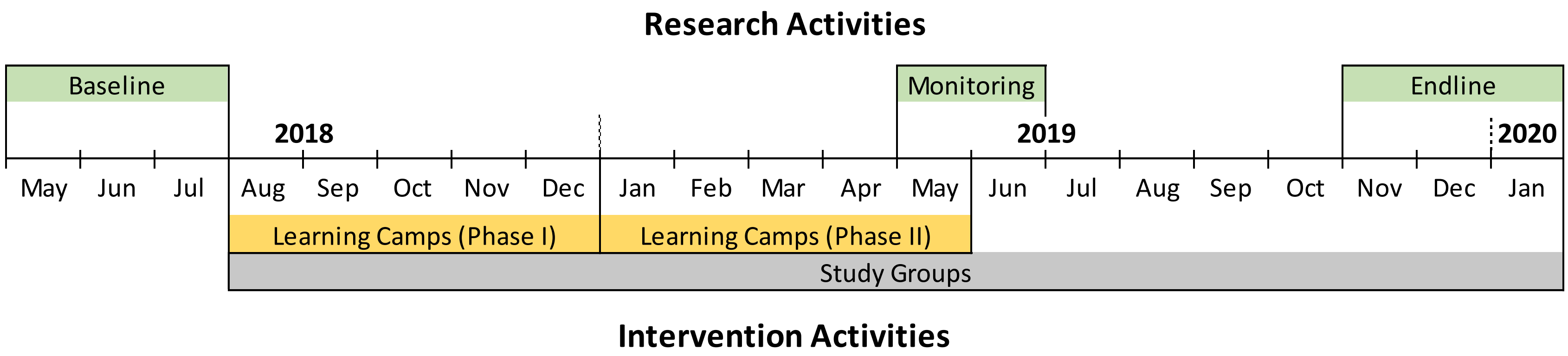


Table 1. Summary statistics across treatment groups at baseline

	C	SG & LC	SG	LC	p-value
Panel A: Schools (N= 200)					
Number of classrooms	2.26	2.00	2.06	2.14	0.79
	(1.12)	(1.73)	(1.39)	(1.11)	
Total teaching staff	4.12	3.70	3.58	3.98	0.64
	(2.56)	(3.29)	(2.07)	(2.14)	
Total enrollment	55.72	52.52	52.92	52.60	0.83
	(19.85)	(20.57)	(23.42)	(20.73)	
Share of girls enrolled	0.51	0.49	0.51	0.50	0.68
	(0.11)	(0.07)	(0.08)	(0.07)	
Panel B: Children (N= 5726)					
Age	7.64	7.61	7.61	7.69	0.77
	(1.55)	(1.53)	(1.56)	(1.59)	
Girl	0.49	0.50	0.50	0.51	0.80
	(0.50)	(0.50)	(0.50)	(0.50)	
Present in school	0.75	0.74	0.68	0.68	0.35
	(0.44)	(0.44)	(0.47)	(0.47)	
Likes going to school [1-5]	4.54	4.57	4.59	4.51	0.33
	(0.85)	(0.81)	(0.78)	(0.89)	
Study outside school	0.75	0.74	0.76	0.71	0.25
	(0.43)	(0.44)	(0.43)	(0.45)	
ASER score (language)	1.72	1.66	1.63	1.75	0.78
	(1.43)	(1.42)	(1.41)	(1.44)	
ASER score (math)	1.43	1.40	1.35	1.42	0.67
	(0.85)	(0.89)	(0.85)	(0.84)	
Tracked at endline (share)	0.93	0.92	0.93	0.94	0.59
	(0.25)	(0.28)	(0.25)	(0.24)	
Panel C: Primary Caregivers (N= 4592)					
# HH members	5.15	5.21	5.30	5.24	0.55
	(1.65)	(1.70)	(1.82)	(1.76)	
# children enrolled	1.42	1.47	1.43	1.44	0.61
	(0.78)	(0.82)	(0.84)	(0.82)	
Asset Index	0.19	-0.13	0.04	-0.10	0.23
	(1.78)	(1.81)	(1.83)	(1.89)	
Primary caregiver is literate	0.53	0.51	0.50	0.48	0.73
	(0.50)	(0.50)	(0.50)	(0.50)	
Would like child to go to university	0.45	0.45	0.49	0.44	0.61
	(0.50)	(0.50)	(0.50)	(0.50)	
Pays tuitions	0.24	0.19	0.21	0.19	0.27
	(0.43)	(0.40)	(0.41)	(0.39)	
Estimates language level correctly	0.34	0.30	0.33	0.33	0.37
	(0.47)	(0.46)	(0.47)	(0.47)	
Overestimates language level	0.50	0.51	0.52	0.50	0.87
	(0.50)	(0.50)	(0.50)	(0.50)	
Estimates math level correctly	0.37	0.38	0.36	0.39	0.69
	(0.48)	(0.49)	(0.48)	(0.49)	
Overestimates math level	0.54	0.52	0.56	0.52	0.49
	(0.50)	(0.50)	(0.50)	(0.50)	
Tracked at endline (share)	0.93	0.93	0.94	0.95	0.38
	(0.25)	(0.26)	(0.24)	(0.22)	

Notes: Standard deviations shown in parentheses. The last column shows the p-value from testing whether the mean is equal across all treatment groups (H_0 := mean is equal across groups).

Table 2. Program implementation

Panel A: Children		Learning Camps			Study Groups		Interaction with Pratham		
	Heard of Pratham	Pratham Tested	Pratham Teachers	Divided in groups	in village	participated	Broad definition	LC	SG
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LC & SG	0.16***	0.13***	0.10***	0.35***	0.24***	0.21***	0.36***	0.36***	0.24***
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
SG	0.05**	0.03**	0.01	0.16***	0.16***	0.14***	0.21***	0.16***	0.16***
	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LC	0.10***	0.08***	0.06***	0.28***	0.01	0.01	0.27***	0.29***	0.01
	(0.02)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.03)	(0.01)
Mean control group	0.08	0.05	0.03	0.24	0.07	0.03	0.34	0.26	0.07
Observations	5,328	5,328	5,328	5,328	5,328	5,328	5,328	5,328	5,328

Panel B: Caregivers	Pratham			Learning Camps			Study	Interaction with Pratham	
	Heard of	Interacted with	Material	In school	TL activities	Diff. gr.	Groups	Broad	LC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LC & SG	0.15***	0.09***	0.36***	0.12***	0.10***	0.23***	0.26***	0.37***	0.28***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
SG	0.09***	0.06***	0.25***	0.08***	0.06***	0.14***	0.23***	0.24***	0.17***
	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
LC	0.08***	0.03**	0.24***	0.05***	0.04***	0.13***	0.04***	0.26***	0.15***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Mean control group	0.07	0.02	0.10	0.02	0.02	0.05	0.02	0.14	0.07
Observations	4,251	4,224	4,265	4,251	4,251	4,265	4,265	4,265	4,265

Notes: The dependent variables in Panel A are taken from the children's surveys and are indicators capturing whether: child ever heard of Pratham (1); Pratham performed any test to assess learning levels (2); any Pratham staff replaced the regular teacher (3); students were divided in groups, based on learning level, and combined with students from other grades (4); Study Groups are arranged in the village after school (5); respondent participated in these groups (6); respondent said yes to any question related to Pratham's activities (7); respondent said yes to any question related to Learning Camps activities (8); respondent said yes to any question related to Study Groups activities (9). The dependent variables in Panel B are taken from the caregivers' surveys and are indicators capturing whether: respondent ever heard of Pratham (1); respondent ever interacted with Pratham (2); respondent ever saw Pratham's teaching-learning material (3); Pratham worked in the school (4); Pratham placed the child in a different grade than his/her usual (5); the child worked in groups different than his/her usual class (6); any Study Groups was arranged in the village (7); respondent said yes to any question related to Pratham's activities (8) respondent said yes to any question related to Learning Camps activities (9). LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the full program (which included both learning camps and community Study Groups), to the study groups alone, or to the learning camps alone, respectively. All regressions include students' gender, age, and grade fixed effects. The varying number of observations in Panel B is due to some respondents not knowing the answer to some specific questions. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 3. Program Impact on Test Scores

Dep Var:	Math		Language	
	ASER	Grade II	ASER	Grade II
LC & SG	0.11**	0.06**	0.12**	0.04*
	(0.05)	(0.03)	(0.05)	(0.02)
SG	-0.01	-0.00	0.01	0.02
	(0.05)	(0.03)	(0.06)	(0.02)
LC	0.02	0.02	0.03	0.00
	(0.04)	(0.03)	(0.06)	(0.02)
Basic Controls	✓	✓	✓	✓
Mean control group	1.95	0.30	2.44	0.31
Observations	5,328	5,328	5,328	5,328
No. of clusters	200	200	200	200
p-val(LC & SG=SG)	0.03	0.06	0.07	0.40
p-val(LC & SG=LC)	0.08	0.15	0.10	0.13

Notes: The dependent variables are children's ASER test score (columns 1 and 3), and an indicator for the student reaching at least Grade II level in the ASER test (columns 2 and 4). LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the full program (which included both learning camps and community Study Groups), to the study groups alone, or to the learning camps alone, respectively. *Basic controls* include age, gender, baseline value of the dependent variable and grade fixed effects. The p-values in the bottom of the table are the test of the null hypothesis of equal treatment effects between different intervention arms. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 4. Program impact on children's outcomes

Dep Var:	Attendance		Study outside				ASTE		
	Child	Careg.	Overall	in Groups	Study up to...	Educ. important	Likes school	School perc.	Psych. wellb.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LC & SG	0.018	-0.029	0.065***	0.027***	-0.056	0.205**	-0.024	-0.016	-0.024
	(0.034)	(0.051)	(0.022)	(0.008)	(0.191)	(0.084)	(0.042)	(0.038)	(0.035)
SG	-0.002	-0.006	0.049**	0.012**	-0.175	0.207**	-0.091*	-0.061	-0.009
	(0.041)	(0.050)	(0.023)	(0.005)	(0.165)	(0.081)	(0.048)	(0.043)	(0.033)
LC	0.014	-0.044	0.037	-0.002	-0.250	0.080	-0.016	-0.020	0.009
	(0.035)	(0.057)	(0.023)	(0.002)	(0.187)	(0.087)	(0.039)	(0.039)	(0.033)
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean control group	0.75	5.07	0.77	0.00	11.26	9.13			
R-squared	0.003	0.503	0.008	0.012	0.105	0.021			
Observations	5,328	3,375	5,328	5,328	4,329	5,328	5,320	5,316	5,328
No. of clusters	200	200	200	200	200	200			
p-val(LC & SG=SG)	0.63	0.62	0.45	0.12	0.53	0.99			
p-val(LC & SG=LC)	0.92	0.78	0.18	0.00	0.35	0.14			

Notes: The dependent variables are defined as follows: child at endline was surveyed in school (1); average number of school days attended in the past week (from caregiver's survey) (2); child studies outside school (3); child meets in study groups to study with friends (4); child would like to study until grade X (5); importance the child attribute to education for the future (1-10 scale) (6); average standardized treatment effect (ASTE) across 3 variables on child's enjoyment with school, reading, and mathematics (7); ASTE across 8 variables on child's perceptions of school and learning (8); ASTE across 16 variables on child's psychological well-being (9). LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the full program (which included both learning camps and community Study Groups), to the study groups alone, or to the learning camps alone, respectively. *Basic controls* include age, gender, baseline value of the dependent variable (except for ASTE regressions) and grade fixed effects. The p-values in the bottom of the table are the test of the null hypothesis of equal treatment effects between different intervention arms. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 5. Program impact on caregivers' outcomes

Dep Var:	Involved in SG	Time spent w. child (%)	Help with HW (hr/w)	Total educ. spending	School Int. (ASTE)	Expect. vs actual	
						Lang.	Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LC & SG	0.031***	-0.003	0.117	0.131	0.088**	0.048	0.055
	(0.008)	(0.009)	(0.232)	(0.333)	(0.035)	(0.073)	(0.050)
SG	0.022***	-0.005	0.170	0.022	0.097***	-0.018	0.007
	(0.006)	(0.008)	(0.291)	(0.354)	(0.037)	(0.075)	(0.054)
LC	0.005	0.001	0.202	-0.082	0.079**	-0.072	-0.007
	(0.005)	(0.008)	(0.238)	(0.349)	(0.033)	(0.077)	(0.054)
Basic Controls	✓	✓	✓	✓	✓	✓	✓
Mean control group	0.00	0.24	5.29	3.18		-0.48	-0.69
R-squared	0.010	0.069	0.000	0.005		0.041	0.017
Observations	4,265	4,251	4,250	4,265	4,265	4,157	4,158
No. of clusters	200	200	200	200		200	200
p-val(LC & SG=SG)	0.33	0.78	0.84	0.68		0.40	0.35
p-val(LC & SG=LC)	0.00	0.64	0.68	0.40		0.13	0.22

Notes: The dependent variables are defined as follows: respondent is involved in an after-school study group (1); share of time in an average day spent studying, playing, or reading with the child (2); hours per week spent helping child with homework (3); total amount spent on education-related items for the child (4); average standardized treatment effect (ASTE) across 5 variables on the caregiver's interaction with the school (5); gap between child's ASER test performance and caregiver's expectation (language) (6); gap between child's ASER test performance and caregiver's expectation, (mathematics) (7). LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the full program (which included both learning camps and community Study Groups), to the study groups alone, or to the learning camps alone, respectively. *Basic controls* include age and gender of the child, baseline value of the dependent variable (except for ASTE regressions) and grade fixed effects. The varying number of observations is due to some respondents not knowing the answer to some specific questions. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 6. Program impact on schools' outcomes

Dep Var:	Teachers		Investments since August 2018					Interactions with			
	Total (#)	Present (%)	Construc.	Repair	TLM	Mats	ASTE	SMC	PTA	MoE	ASTE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
LC & SG	-0.360	0.010	-0.045	0.142	-0.167*	-0.198**	-0.137	-0.198**	-0.165	-0.152**	-0.483***
	(0.318)	(0.047)	(0.057)	(0.101)	(0.101)	(0.080)	(0.096)	(0.087)	(0.106)	(0.072)	(0.157)
SG	-0.060	0.024	-0.000	-0.032	-0.195*	-0.114	-0.160	-0.043	0.043	-0.115	-0.208
	(0.238)	(0.049)	(0.060)	(0.102)	(0.100)	(0.086)	(0.106)	(0.079)	(0.104)	(0.075)	(0.141)
LC	0.115	0.057	-0.067	-0.110	-0.103	-0.088	-0.213**	-0.219**	-0.116	-0.140*	-0.449***
	(0.224)	(0.043)	(0.050)	(0.101)	(0.100)	(0.088)	(0.097)	(0.088)	(0.103)	(0.072)	(0.154)
Baseline Y	✓	✓	✓	✓	✓	✓		✓	✓	✓	
Mean control group	3.78	0.79	0.10	0.52	0.56	0.30		0.84	0.56	0.90	
R-squared	0.633	0.040	0.046	0.035	0.041	0.033		0.046	0.028	0.134	
Observations	199	199	195	196	196	195	199	199	186	197	199
No. of clusters	0.19	0.76	0.42	0.09	0.78	0.25		0.09	0.05	0.67	
p-val(LC & SG=SG)	0.09	0.22	0.65	0.01	0.53	0.15		0.83	0.64	0.89	

Notes: The dependent variables are defined as follows: number of teachers appointed to the school (1); share of teachers present at the school on the day of the survey (2); any investment undertaken in school in the previous 14 months in: construction (3); repair (4); purchase of teaching-learning material (5); purchase of sitting mats for children (6); average standardized treatment effect (ASTE) over the four outcomes in columns 3-6 (7); any School Management Committee meeting in the previous 2 months (8); any parent-teacher meeting in the previous 2 months (9); any visit from the Ministry of Education in the previous 3 months (10); average standardized treatment effect (ASTE) over the three outcomes in columns 8-10. LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the full program (which included both learning camps and community Study Groups), to the study groups alone, or to the learning camps alone, respectively. *Baseline Y* includes the baseline value of the outcome variable. Robust standard errors in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 7. Program Impact across Implementation Phases

Dep Var:	Math		Language	
	ASER	Grade II	ASER	Grade II
	(1)	(2)	(3)	(4)
LCI & SG	0.15**	0.07*	0.19***	0.07*
	(0.06)	(0.04)	(0.07)	(0.03)
LCII & SG	0.07	0.04	0.06	0.02
	(0.06)	(0.03)	(0.06)	(0.03)
SG	-0.01	-0.00	0.01	0.02
	(0.05)	(0.03)	(0.06)	(0.02)
LC Phase I	0.05	0.03	0.07	0.03
	(0.06)	(0.03)	(0.08)	(0.03)
LC Phase II	-0.01	-0.00	-0.01	-0.02
	(0.05)	(0.03)	(0.06)	(0.03)
Basic Controls	✓	✓	✓	✓
Mean control group	1.95	0.30	2.44	0.31
R-squared	0.416	0.214	0.580	0.316
Observations	5,328	5,328	5,328	5,328
No. of clusters	200	200	200	200

Notes: The dependent variables are children's ASER test score (columns 1 and 3), and an indicator for the student reaching at least Grade II level in the ASER test (columns 2 and 4). LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the standard Pratham program (which included both learning camps and community Study Groups), to the commmunity study groups alone, or to the learning camps alone, respectively. "I" and "II" indicate whether the Learning Camps took place during the first or the second implementation phase. *Basic controls* include age and gender of the child, baseline value of the dependent variable and grade fixed effects. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 8. Program Exposure

	Math			Reading		
	LC part. (%)	Density SGs	Both	LC part. (%)	Density SGs	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS Estimates						
LC & SG	0.25***	1.21***	2.25***	0.27***	1.00***	1.04***
	(0.06)	(0.36)	(0.44)	(0.07)	(0.37)	(0.34)
SG	0.03	0.43	0.71*	0.06	0.17	0.22
	(0.05)	(0.41)	(0.40)	(0.06)	(0.48)	(0.48)
LC	0.16***	0.05	0.17***	0.20***	0.02	0.05
	(0.06)	(0.04)	(0.06)	(0.07)	(0.05)	(0.05)
Panel B: 2SLS Estimates						
LC & SG	0.17**	0.97**	1.49**	0.19**	1.13**	1.72**
	(0.07)	(0.41)	(0.63)	(0.08)	(0.49)	(0.75)
SG	0.04	-0.10	-0.10	0.04	0.10	0.09
	(0.07)	(0.57)	(0.57)	(0.09)	(0.70)	(0.70)
LC	-0.01	0.02	0.04	0.01	0.03	0.04
	(0.05)	(0.04)	(0.07)	(0.06)	(0.06)	(0.09)
F-stat. LC & SG	514.73	119.31	55.97	521.03	120.61	56.52
F-stat. LG		125.70	84.30		126.08	84.05
F-stat. LC	846.45		564.72	844.97		563.61
Basic Controls	✓	✓	✓	✓	✓	✓
Mean control group	1.95	1.95	1.95	2.44	2.44	2.44
Observations	5,328	5,328	5,328	5,328	5,328	5,328
No. of clusters	200	200	200	200	200	200

Notes: In columns 1, 3, 4, and 6, Learning Camps exposure is defined as the share of days participating in the camps over the total number of possible days (30). The average exposure among children in the Learning Camps study arms is 68%. In columns 2, 3, 5, and 6, Study Groups exposure is proxied with the number of Study Groups organised in the village, weighted by the number of children enrolled in the local primary school at baseline. The average value of this variable across villages in the Study Groups study arm is 0.1 (i.e. 1 study group every 10 children). *Basic controls* include age and gender of the child, baseline value of the dependent variable and grade fixed effects. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 9. Cost Effectiveness Analysis

	LC & SG	**LC**	**SG**
Total costs (yearly)	**2,342,549**	**2,171,751**	**1,357,756**
Personnel	1,619,820	1,619,820	909,120
TLM	367,200	236,200	232,400
Training	179,091	139,293	39,798
Travel	42,000	42,000	42,000
Other Costs	134,438	134,438	134,438
# years	1.33	1.33	1.33
# villages served	50	50	50
# children served per village	53	53	53
Avg cost per student	1,176	1,090	681
Avg improvement in learning (in SD)	0.09 - 0.12	0.018 - 0.023	-0.01 - 0.01
Cost per 0.1 SD gain	980 - 1,306		
Cost per 0.1 SD gain (USD)	13.7 - 18.3		
Additional SD per 100 USD	0.55 - 0.73		

Appendix A – Additional Figures and Tables

Figure A.1 Distribution of ASER math levels by grade at baseline

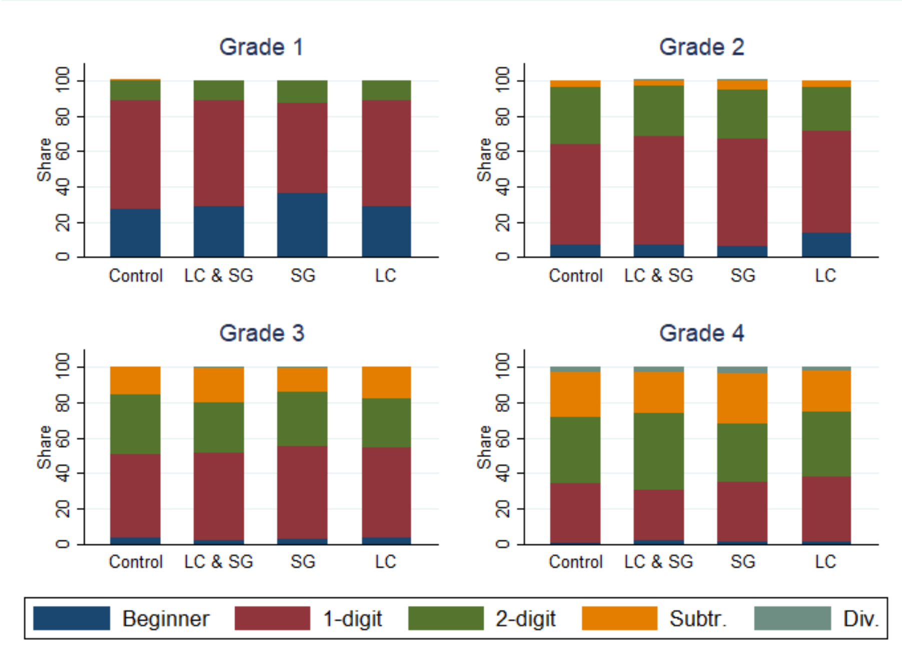


Figure A.2 Distribution of ASER language levels by grade at baseline

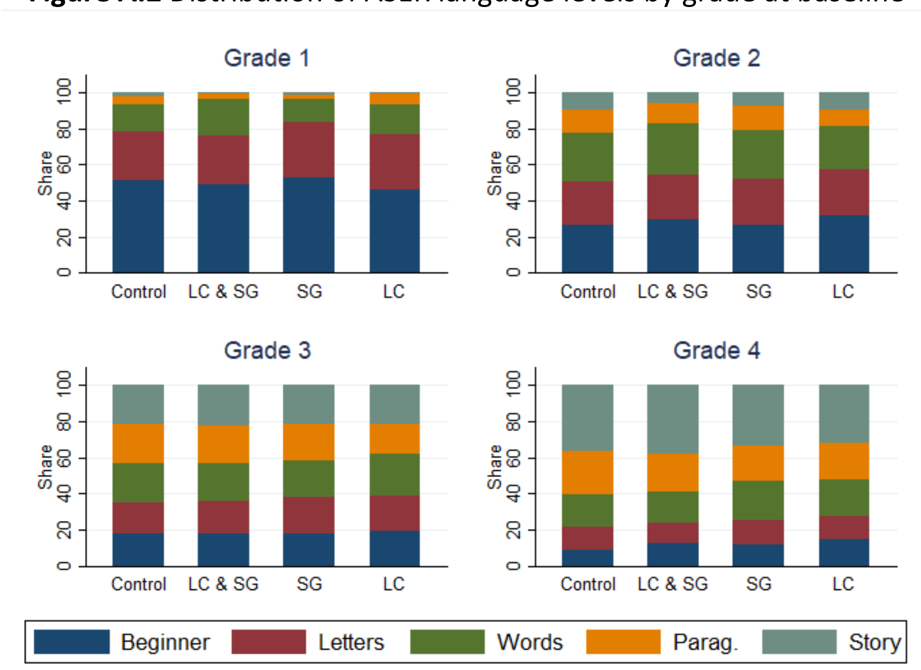
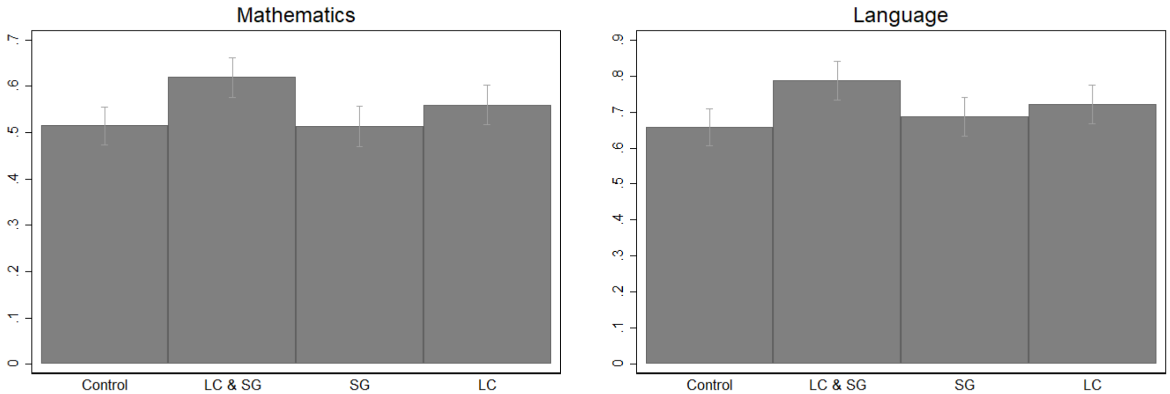


Figure A.3 Learning gains in mathematics and language over the study period



Notes: The figures show the difference in learning level in mathematics (left) and language (right), between endline and baseline, across the entire sample, by treatment arm.

Figure A.4 Distribution of ASER math levels by grade at endline

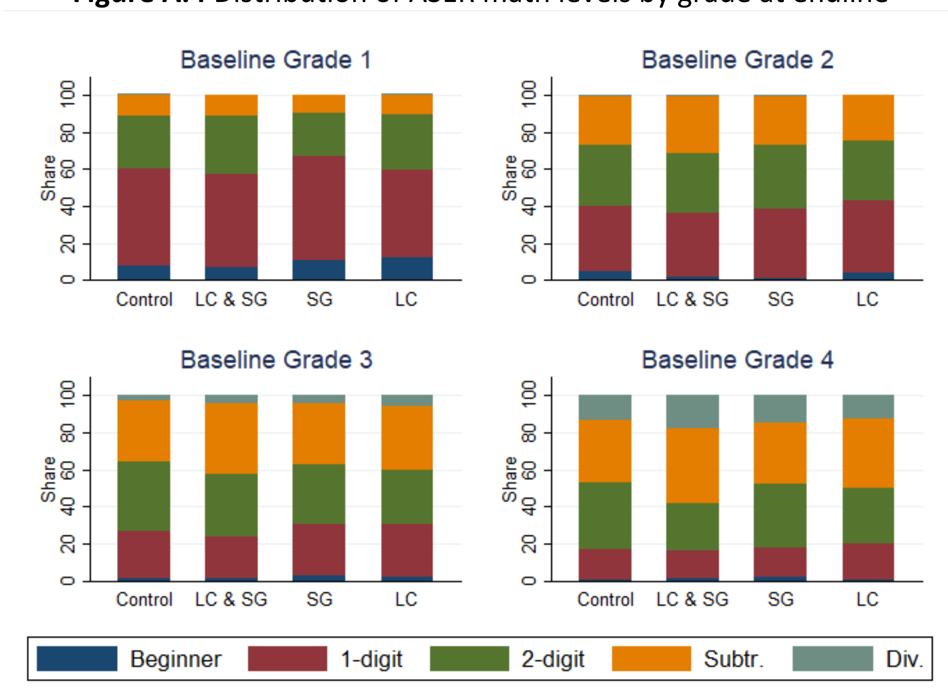


Figure A.5 Distribution of ASER language levels by grade at endline

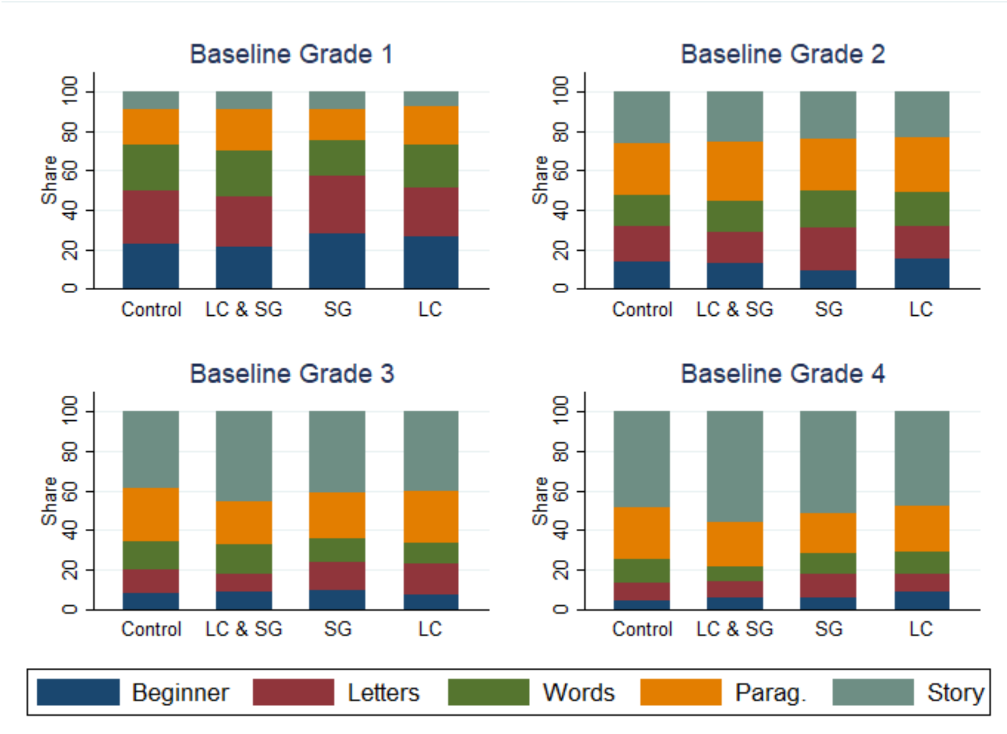
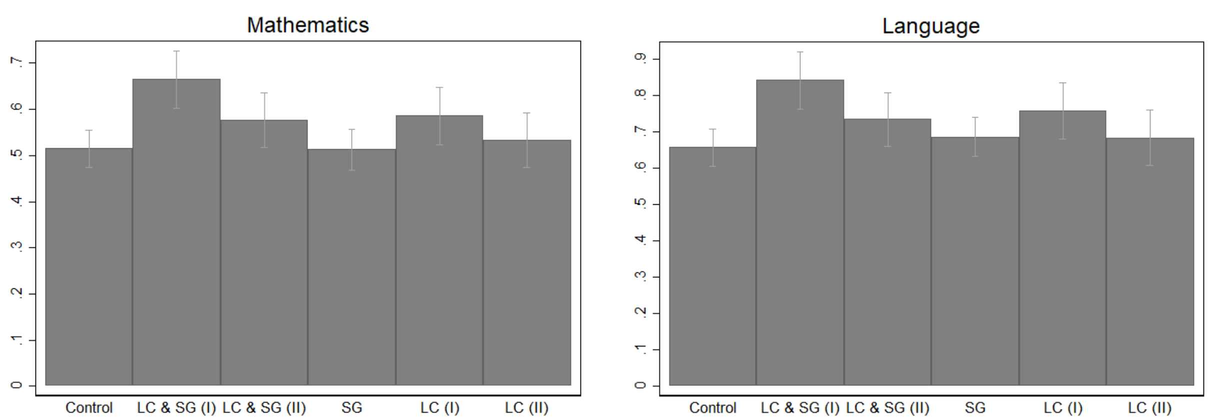


Figure A.6 Learning gains in mathematics and language over the study period, by phase



Notes: The figures show the difference in learning level in mathematics (left) and language (right), between endline and baseline, across the entire sample, by treatment arm. The figure distinguishes between villages that received the learning camps in the first implementation phase (I) and those that received them in the second one (II).

Figure A.7 Cost effectiveness of other education programs evaluated with RCTs

Source: Bhula et al (2013)

Table A.1 Sample and Attrition by Grade

	Child survey				Caregiver survey			
Grade at baseline	Baseline	Endline	Attrition	Attrition (%)	Baseline	Endline	Attrition	Attrition (%)
Grade 1	1449	1334	115	0.079	1168	1091	77	0.066
Grade 2	1373	1277	96	0.070	1106	1032	74	0.067
Grade 3	1460	1349	111	0.076	1160	1086	74	0.064
Grade 4	1444	1368	76	0.053	1158	1094	64	0.055

Table A.2 Attrition Checks

Dep Var:	Attrited Child					
Interaction with...:			Grade	Girl	ASER	
					Math	Language
	(1)	(2)	(3)	(4)	(5)	(6)
LC & SG	0.004	0.006	0.016	0.000	0.002	-0.001
	(0.014)	(0.014)	(0.028)	(0.018)	(0.027)	(0.021)
SG	0.019	0.021	0.038	0.018	0.025	0.033
	(0.015)	(0.015)	(0.028)	(0.018)	(0.027)	(0.024)
LC	0.007	0.009	-0.005	0.012	-0.005	-0.002
	(0.015)	(0.014)	(0.025)	(0.019)	(0.024)	(0.020)
LC & SG × ...			-0.004	0.013	0.004	0.004
			(0.009)	(0.019)	(0.013)	(0.006)
SG × ...			-0.007	0.007	-0.003	-0.008
			(0.009)	(0.017)	(0.012)	(0.008)
LC × ...			0.005	-0.007	0.009	0.006
			(0.008)	(0.018)	(0.011)	(0.006)
Basic Controls	✗	✓	✓	✓	✓	✓
Mean control group	0.06	0.06	0.06	0.06	0.06	0.06
R-squared	0.001	0.017	0.017	0.017	0.022	0.025
Observations	5,726	5,726	5,726	5,726	5,726	5,726
No. of clusters	200	200	200	200	200	200

Notes: LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the standard Pratham program (which included both learning camps and community Study Groups), to the commmunity study groups alone, or to the learning camps alone, respectively. All regressions with interactions include interaction components as well (not reported). *Basic controls* include: grade fixed effects, gender, and age of the child. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level;, ***Significant at 1% level.

Table A.3 Program Impact using Test A

Dep Var:	ASER test scores							
Model:	OLS		OLS		OLS		2SLS	
Subject:	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LC & SG	0.03**	0.01						
	(0.01)	(0.01)						
SG	0.00	-0.01	0.00	-0.01				
	(0.01)	(0.01)	(0.01)	(0.01)				
LC	0.00	-0.00						
	(0.01)	(0.01)						
LCI & SG			0.04***	0.01				
			(0.01)	(0.01)				
LCII & SG			0.01	0.00				
			(0.02)	(0.01)				
LC Phase I			0.01	-0.00				
			(0.02)	(0.01)				
LC Phase II			-0.00	-0.00				
			(0.01)	(0.01)				
LC (%) & SG (density)					0.42***	0.10	0.36**	0.11
					(0.13)	(0.08)	(0.17)	(0.12)
SG (density)					0.19	-0.08	0.04	-0.12
					(0.13)	(0.08)	(0.16)	(0.11)
LC (part. %)					0.03*	0.02	0.01	-0.00
					(0.02)	(0.01)	(0.02)	(0.01)
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Mean control group	0.63	0.76	0.63	0.76	0.63	0.76	0.63	0.76
R-squared	0.459	0.518	0.460	0.518	0.461	0.518	0.460	0.517
Observations	5,328	5,328	5,328	5,328	5,328	5,328	5,328	5,328
No. of clusters	200	200	200	200	200	200	200	200

Notes: The dependent variables are children's share of correct answers in the math and language components of "Test A". LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the full program (which included both learning camps and community Study Groups), to the study groups alone, or to the learning camps alone, respectively. In columns 5 to 8, Learning Camps exposure is defined as the share of days participating in the camps over the total number of possible days (average among children in the Learning Camps study arms is 68%), while Study Groups exposure is proxied with the number of Study Groups organised in the village, weighted by the number of children enrolled in the local primary school at baseline (average across villages in the Study Groups study arm is 0.1 – i.e. 1 study group every 10 children). *Basic controls* include age and gender of the child, baseline value of the dependent variable and grade fixed effects. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table A.4 Heterogneity Analysis

	...Girl		...Grade 3–4		...BL Score		...Sch. BL < Med.	
	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LC & SG × ...	-0.00	-0.12	-0.02	0.16**	0.00	-0.01	-0.01	0.13
	(0.07)	(0.08)	(0.05)	(0.08)	(0.04)	(0.03)	(0.08)	(0.11)
SG × ...	0.02	-0.09	0.06	0.12	0.01	0.04	-0.04	-0.04
	(0.06)	(0.08)	(0.06)	(0.08)	(0.04)	(0.04)	(0.08)	(0.11)
LC × ...	-0.01	-0.04	0.03	0.12	0.03	0.03	0.07	0.16
	(0.07)	(0.07)	(0.05)	(0.07)	(0.03)	(0.03)	(0.08)	(0.11)
LC & SG	0.11*	0.18**	0.12**	0.05	0.10	0.14	0.13**	0.08
	(0.06)	(0.07)	(0.05)	(0.06)	(0.07)	(0.10)	(0.05)	(0.06)
SG	-0.02	0.06	-0.04	-0.05	-0.03	-0.07	0.06	0.07
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.11)	(0.06)	(0.06)
LC	0.03	0.05	0.01	-0.03	-0.02	-0.02	0.03	-0.02
	(0.06)	(0.07)	(0.05)	(0.07)	(0.07)	(0.10)	(0.06)	(0.08)
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓
Mean control group	1.95	2.44	1.95	2.44	1.95	2.44	1.95	2.44
R-squared	0.415	0.580	0.415	0.580	0.415	0.580	0.432	0.584
Observations	5,328	5,328	5,328	5,328	5,328	5,328	5,328	5,328
No. of clusters	200	200	200	200	200	200	200	200

Notes: LC & SG, SG and LC are indicator variables that take on value 1 if the child's village was assigned to the standard Pratham program (which included both learning camps and community Study Groups), to the commmunity study groups alone, or to the learning camps alone, respectively. In columns 7 and 8 they are interacted with an indicator that takes on value 1 if the weighted average (by grade) of the baseline students' test scores is below the median across the sample. *Basic controls* include age and gender of the child, baseline value of the dependent variable and grade fixed effects. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Appendix B – Test Administration

The tests were administered individually to each child by trained enumerators, in local (Assamese) language. The content was based on well-established assessment tools that have already been extensively piloted and used across India. As mentioned in the main text, two tests were used in this project.

The first test mirrored the standard ASER (Annual Status of Education Report) test, a nationwide test conducted yearly by the ASER Center all over India for children aged 5 to 16. The test is divided in a math and a language component and had already been extensively piloted, tested, and used in previous evaluations conducted in India, most recently by Banerji et al (2017). According to the ASER test, children are classified in five categories based on their knowledge. For mathematics the categories are:

- Beginner (no number recognition)
- single-digit number recognition
- double-digit number recognition
- subtraction (of double-digit numbers)
- division (of a double-digit number by a single digit)

For language the categories are:

- Beginner (cannot recognize letters),
- recognize letters
- reads words
- reads a paragraph
- reads a short story

Figure B.1 shows an English example of an ASER test.

The second test, “test A”, was instead created by the research team, based on tests that had been previously used in other studies conducted in India (Muralidharan et al. 2018). The test, which contained 13 questions, was especially designed to target children in the lowest grades and contained basic math and language questions. Figure B.2 shows a sample of questions included in the test. By endline, test A resulted too easy for the students in the sample and we faced severe top-coding issues. For this reason, in our main analysis we excluded this test, although results for the math section (which had lower top-coding issues) are very consistent.

Figure B.1 ASER Test – sample questions

READING TEST SAMPLE (1)

Story

Salma is a little girl. She had a pretty doll. She loved playing with her doll. One day the doll fell from her hand to the floor. It broke into many pieces. Salma was very sad. She cried a lot. Her mother gave her another doll. Now she is happy again.

Para

Ravi is a boy.
He has many friends.
He loves to draw.
He does not like to sing.

Para

My village is very big.
It has many houses.
It also has a shop.
The bus stops in my village.

READING TEST SAMPLE (1)

Letter

b s o
k m
y r h
t x

Ask the child to recognize any 5 letters. At least 4 must be correct.

Word

ring bad
ball
cold king
clap foot
fan
girl crow

Ask the child to read any 5 words. At least 4 must be correct.

MATH TEST SAMPLE (1)

Number recognition 1—9	Number recognition 10—99	Subtraction	Division
1 4	51 83	46 − 29 63 − 39	7)879
7 3	37 65	47 − 28 45 − 17	6)824
6 9	55 26	92 − 76 84 − 57	8)985
5 2	91 43	52 − 14 66 − 48	4)517
	36 27		
Ask the child to recognize any 5 numbers. At least 4 must be correct.	Ask the child to recognize any 5 numbers. At least 4 must be correct.	Ask the child to do any 2 subtraction problems. Both must be correct.	Ask the child to do any 1 division problem. It must be correct.

Figure B.2 Test A – sample questions