Journal of Development Economics 184 (2027) 103906



Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec



Regular article



Supporting learning in and out of school: Experimental evidence from India☆

Martina Björkman Nyqvist [a,b,c,d,e,*], Andrea Guariso [f,g,h,b]

[a] Stockholm School of Economics, Sweden
[b] Misum, Sweden
[c] J-PAL, United States of America
[d] CEPR, United Kingdom
[e] BREAD, United Kingdom
[f] University of Milan-Bicocca, Italy
[g] TIME, Ireland
[h] INSPIRE-LEAP, Italy

ARTICLE INFO

Dataset link: https://doi.org/10.17632/mtf2x29p69.1

Keywords:
Primary school
India
Learning
Study groups
Pedagogic intervention

ABSTRACT

Targeted educational interventions for primary school children in low-income countries have led to substantial improvements in learning, but their impact varies considerably across studies, raising questions about how to sustain their effectiveness in different settings. In this paper, we study a novel primary education program that combines a conventional in-school targeted pedagogical intervention with an out-of-school study group initiative facilitated by the community. To identify the impact of the program and its components, we design a cross-cutting randomized experiment across 200 villages in Assam, India. Our results show that standalone implementation of either component has no significant effect on learning outcomes, but when the components are combined in the full program, they improve children's test scores in mathematics and language by 0.11 and 0.09 standard deviations, respectively. Our results suggest that integrating in-school targeted instruction with community-based out-of-school activities can enhance program effectiveness by increasing students' engagement and overall exposure to learning, at relatively low cost.

1. Introduction

The dramatic gains in school enrollments that have been achieved in low- and middle-income countries over the past two decades have not been accompanied by comparable improvements in learning outcomes (World Bank, 2018; Altinok et al., 2018; Le Nestour et al., 2022). India is a case in point: despite achieving nearly universal enrollment for children between the ages of 6 and 14, the percentage of grade 5 students capable of reading a grade 2 text or solving a grade 2 subtraction problem has stagnated below 50% for more than ten years (ASER, 2018). The COVID-19 pandemic and the associated school closures have further slowed learning (Moscoviz and Evans, 2022; Guariso and Björkman Nyqvist, 2023; Singh et al., 2024), making the quest for effective approaches to boost students' achievements an even more pressing priority. The existing evidence primarily focuses on approaches to make schools more effective and the time spent in school more productive (see Muralidharan 2017 for a review). Targeted pedagogical interventions that are tailored to the specific needs of the students have gained popularity as one of the most effective

☆ We thank the editor and two anonymous referees for their valuable comments and suggestions, which greatly improved the paper. We are grateful for comments and suggestions by Andy de Barros, Timothée Demont, Paul Glewwe, Selim Gulesci, Isaac Mbiti, Abhijeet Singh, Gregory F. Veramendi, and seminar participants at University of Munich (LMU), UC Merced, University of Goettingen, Aix-Marseille School of Economics (AMSE), University of Bologna, University of Milan-Bicocca, Trinity College Dublin, MWIEDC, EOS Development Webinar, YoDEV Webinar, as well as Léo Komminoth, Fadhil Nadhif Muharam, Artur Obminski, Merve Demirel, Ranjana Sinha, 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 Banerji and Saveri Kulshreshth at Pratham India and Ingrid Eelde Koivisto at Pratham Sweden for insightful discussions about the programs and for their collaboration throughout the study, as well as for sharing administrative data. All mistakes are our own. Financial support from Carl Bennet AB, the Swedish Research Council (2016-05615) and the Mistra Foundation is greatly appreciated. Andrea also acknowledges funding from the TCD Arts and Social Sciences Benefactions (ASSB) Fund. The study was registered with the AEA RCT registry (0002817) and received ethical approval from IFMR Human Subjects Committee (IRB00007107) and TCD Ethics Review Board (05062018).

* Corresponding author at: Stockholm School of Economics, Sweden.
E-mail address: martina.bjorkman.nyqvist@hhs.se (M. Björkman Nyqvist).

https://doi.org/10.1016/j.jdeveco.2026.103906
Received 29 October 2024; Received in revised form 3 July 2026; Accepted 7 August 2026
Available online 21 August 2026
0304-3878/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

approaches to enhancing students' learning (e.g. Banerjee et al., 2007, 2016; Glewwe and Muralidharan, 2016; Muralidharan et al., 2019). However, their effectiveness has been shown to vary considerably across studies (e.g. Angrist and Meager, 2023; Cabezas et al., 2021; Duflo et al., 2024), raising questions on how to best ensure and preserve the effectiveness of these models across time and space. One common feature of these pedagogical interventions is that, by targeting instruction, they focus on what happens in class, while they do not provide support for after-school practice and learning.

In this paper, we partner with Pratham – one of the largest Indian NGOs – to study whether it is possible to strengthen a standard targeted pedagogical intervention implemented in primary schools with a community-managed study groups program that takes place after school in the village, providing in this way an opportunity for children to meet and study together and consolidate what they learn in class. The in-school pedagogical program is modeled around the well-known "Teaching at the Right Level" (TaRL) approach and is delivered in class through *Learning Camps* sessions for a total of 30 days. During camp days, students are reorganized into groups within the school according to their actual knowledge level, rather than their assigned grade, and receive tailored instruction led by the NGO staff, matched to their level.[1] The out-of-school program instead consists of *Study Groups* managed by local volunteers (typically mothers) and brings together primary school children of different ages to study and practice in small groups using a peer-to-peer approach outside of school hours. These groups are meant to extend learning practices beyond the classroom, with the support of the community.

We study the impact of the program on children's learning using a randomized controlled trial with factorial design, implemented across 200 villages in Assam, northeast India. The villages were randomly assigned to one of four study arms of equal size: full program (with both Learning Camps and Study Groups), Learning Camps only, Study Groups only, and control group. We conducted two rounds of data collection: prior to the start of the program (mid-2018) and approximately 17 months later (end of 2019). We surveyed the primary public schools located in each village (N = 200), a representative sample of children enrolled within each school (N = 5726), and a representative sample of their caregivers (N = 4592).

The program was implemented with high fidelity and in accordance with treatment assignment. Administrative data show that study groups and learning camps took place in all target villages, with the exception of one case where a control school and a full-program school were accidentally swapped due to similar names.[2] Our data show balance across treatment arms at baseline, and we can exclude the presence of attrition bias. Despite high implementation fidelity, the level of program participation and uptake was far from universal — a common challenge across many educational interventions. Administrative data reveal that students in the arms assigned to the learning camps attended, on average, 20 out of 30 camp days, while in the villages assigned to the study groups, on average 5 groups were created in the villages, serving a total of about 30 students.[3] Self-reported data, which provide a lower bound on true program exposure due to imperfect recollection, reveal that at endline less than one third of students in the Learning Camps arms recollect participating in camp activities, and less than one-fifth of students in the Study Groups arms recollect participating in the groups. We find that in villages where both components were implemented, self-reported exposure to the interventions is significantly higher, indicating that the combination of in-school and out-of-school activities increased awareness of the individual programs and total exposure to learning activities.

Our empirical analysis highlights five key results. First, we find that when the two program components are implemented in isolation, neither generates a detectable impact on students' learning.

Second, we find that when the two components are implemented together, the program raises children's learning in mathematics and language on average by 0.11 and 0.09 standard deviations (SD), respectively, compared to children in the control group.[4] These effects translate into a significant increase in the share of children that achieve minimum standards (i.e. grade-2 knowledge) in mathematics and language, by 20% and 13%, respectively.

Third, we explore the heterogeneity of these effects and find suggestive evidence that timing matters: by leveraging random variation in the start of the learning camps across schools, we find that children who were exposed to the learning camps earlier achieve higher test scores by the end of the study. This pattern is consistent with earlier exposure placing children on a steeper learning trajectory, but it may also reflect other features of the implementation schedule, including differences in coordination between in-school and out-of-school components and seasonality effects. We do not find any clear evidence of differential impact across grade, gender, or initial learning level.

Fourth, we study the impact on additional (pre-specified) secondary outcomes and find that none of the interventions impacted the overall amount of time parents spend with their children, but they all led to increased parental engagement with education through more frequent interactions with the school. While increased parental engagement is an encouraging result, as it is a dimension notoriously difficult to move (e.g. Di Maro et al., 2020), the limited impact on learning outcomes highlights the challenge of translating parental engagement into higher learning. At the school level, we find evidence of a substitution effect: schools exposed to the programs, and especially those where the Learning Camps took place, experienced a drop in educational inputs in terms of both physical resources and interactions with stakeholders. This substitution effect is in line with what Das et al. (2013) found in India and Zambia by looking at changes in households' inputs following the disbursement of school grants.

Finally, we show in a cost-effectiveness analysis that the full program costs between 15 and 18.3 US$ per student per 0.1 SD increase in learning, which puts it in the middle of the cost-effectiveness distribution of 27 education programs assessed by J-PAL (Bhula et al., 2013). Our analysis highlights that in the context of the ongoing expansion of the TaRL model across India and Africa, adding the out-of-school component can lead to significant gains in children's learning levels at a low additional cost.

In discussing the pattern of our results – in particular the absence of detectable impacts when the two programs are implemented in isolation – we present evidence suggesting that these findings are unlikely to be driven by poor implementation quality or cross-village spillovers. Instead, they are consistent with the presence of participation constraints, school-level behavioral responses, and context-specific limits to the marginal effectiveness of standalone interventions. Bundling the two components increases the overall dose of instruction and engagement sufficiently to overcome these constraints and generate detectable learning gains.

Our primary contribution relates to the literature on targeted instruction in low-income settings. A large body of evidence identifies

[1] The Learning Camps model studied here builds on the program previously evaluated by Banerjee et al. (2016) and reflects the design adopted by Pratham for its scale-up across India. More details on the comparison between the two models are reported in Section 2.2.

[2] In the analysis, we adopt a conservative approach and rely on the original random assignment, although using actual assignment does not change any of our conclusions.

[3] The average size of the public primary school in the study villages is 53 children.

[4] Throughout the paper we refer to the commonly used ASER test as our preferred measure of learning. When considering an alternative test, which, as explained below, suffers from top-coding issues, results on mathematics are confirmed, while those on language become weaker (see Section 4.1).

pedagogical approaches that adapt instruction to children's learning levels as among the most promising strategies for improving learning outcomes in low- and middle-income countries (e.g. Glewwe and Muralidharan, 2016; McEwan, 2015; Global Education Evidence Advisory Panel, 2020; Muralidharan, 2017). Much of this evidence has been shaped by the TaRL approach pioneered by Pratham in India, which has delivered substantial learning gains in several influential evaluations (e.g. Banerjee et al., 2007, 2016). At the same time, a growing literature emphasizes that the effectiveness of targeted instruction varies considerably across settings and implementation modalities. Angrist and Meager (2023), synthesizing evidence from eight evaluations of TaRL-type programs, document substantial heterogeneity in estimated impacts across study arms, with learning improvements ranging from 0.07 to 0.78 standard deviations. They show that much of this variation is associated with differences in implementation, including take-up, fidelity, and delivery models. The largest effects have been documented in NGO-led and camp-based implementations in India (Banerjee et al., 2007, 2016), as well as in some technology-assisted models (Muralidharan et al., 2019). In contrast, interventions embedded in regular school systems or delivered by volunteers and regular teachers often generate more modest gains (Cabezas et al., 2021; Duflo et al., 2011, 2024). Importantly, however, there remains substantial variation even within similar delivery models, as illustrated in Appendix Table A.1, which summarizes the existing experimental evidence on the impacts of targeted educational interventions. Our study contributes to this literature by showing that combining a standard in-school targeted instruction program with a simple, community-managed after-school intervention generates significant improvements in learning, even in a context where the in-school intervention alone is not effective. More broadly, our results suggest that expanding opportunities for exposure and practice beyond the classroom may be a key margin through which targeted instruction translates into improved learning outcomes.

We also contribute to the literature that investigates the impact of out-of-school learning programs. Despite the potential complementarities between in-school and out-of-school programs, the literature has generally focused on the former and paid little attention to the latter. There are a few studies that look at the impact of after-school remedial education (Lakshminarayana et al., 2013; Banerjee et al., 2016; Chiplunkar et al., 2023; Romero et al., 2022), sometimes incorporating computer-assisted learning (Linden, 2008; Lai et al., 2015; Muralidharan et al., 2019), but most of them are implemented in schools immediately before or after official school hours and are led by teachers or specifically trained personnel. Despite the well-recognized benefit that parental involvement brings to child development in the early years of life (Attanasio et al., 2020a,b), there is little evidence of programs designed to engage parents in the education process of their children in low-income settings, where parents themselves might have little or no experience with formal education. The existing limited evidence has so far struggled to identify effective interventions. Banerji et al. (2017) studied an out-of-school program in India focused on adult literacy and parental involvement, finding only a modest impact (ranging from 0.035 to 0.056 SD) on children's learning. More recently and in line with our study, de Barros et al. (2024) studied a pedagogical intervention in India (teachers' training focused on activity-based instruction) both implemented alone and in combination with an out-of-school component consisting of a community-led student contest, finding no impact on students' learning. Finally, Chimbutane et al. (2026) studied a light-touch literacy-promoting intervention in Mozambique, which included community-led reading camps, finding no impact on learning. Rigorous evidence on out-of-school parental participation programs in high-income countries is also limited. A randomized evaluation of a program in a poor school district in France that aimed to enhance parental involvement in the education of adolescent children found significant positive effects on parental involvement and on student behavior in school, but no effects on test scores (Avvisati et al., 2014). We contribute to this literature by studying a new light-touch program implemented outside of the school environment and fully managed by community members (typically mothers). While we do not find that the program is effective in raising learning levels when implemented in isolation, we show that such out-of-school parental engagement can enhance the effectiveness of in-school educational programs in a cost-effective way.

2. Setting and intervention

2.1. Setting

In India primary education is compulsory, starts at age 6, and lasts for eight grades, divided into lower primary (grades 1 to 5) and upper primary (grades 6 to 8). Progression through primary grades is automatic (Government of India, 2009), and the majority of children (69.1%) is enrolled in public schools. Our study takes place in Nagaon district, which is located in the state of Assam (Fig. 1). In Assam, more than 75% of primary school students are enrolled in public schools, where the school year follows the solar year and runs from January to December (ASER, 2018). School enrollment for children aged 6–14 years is nearly universal, reaching 97.7% and exceeding the Indian average of 95.9%. Yet, learning levels remain alarmingly low: only 34% of grade 5 children can read at a grade 2 level, and a mere 14% can solve simple division problems. These figures are well below the corresponding national averages of 44% and 23%, respectively. In terms of schooling outcomes, Nagaon district's performance aligns with the state average (ASER, 2016).

2.2. Intervention

We study two interventions targeting primary school children, called *Learning Camps* and *Study Groups*, both implemented by the Indian NGO Pratham Education Foundation. In this section, we provide an overview of the key features of each program. For more detailed information, including detailed set-up and structure, we refer to Appendix B.

Learning camps. The *Learning Camps* refer to the targeted pedagogical program implemented by Pratham within schools during regular school hours. The camps consist of intensive bursts of teaching and learning activities focused on building foundational skills in reading and arithmetic. The design of the camps is based on Pratham's pedagogical approach known as "Teaching at the Right Level" (TaRL), which customizes teaching to match each child's individual learning level. Pratham introduced the camps in the target villages at the start of the school term and mobilized community volunteers to assist with the implementation. The camps were then administered in the schools by Pratham staff, with the support of teachers and volunteers. The camps were organized in three 10-day periods, for a total duration of 30 days spread over the course of a five-month school term. To determine the reading and arithmetic proficiency of each child, Pratham tested all students in grades 1 through 5 at the beginning of the first camp day. Based on the results, children were then grouped for the length of the camps according to their skill level. Each camp day had an average duration of 4 h and engaged children in language and mathematics activities, utilizing materials that were customized to suit their respective skill groups. An earlier version of the Learning Camps was evaluated by Banerjee et al. (2016) in Uttar Pradesh. Table 1 compares the earlier model to the version implemented in our setting, which is the version Pratham adopted for scale-up across India. The main difference is the duration: camps were reduced from a total of 50 days to 32 (which includes the initial 2-day demonstrative introduction session to the community). In addition, whereas the original model targeted students in grades 3 to 5, the revised model expanded the curriculum to include structured activities for children in grades 1 and 2.

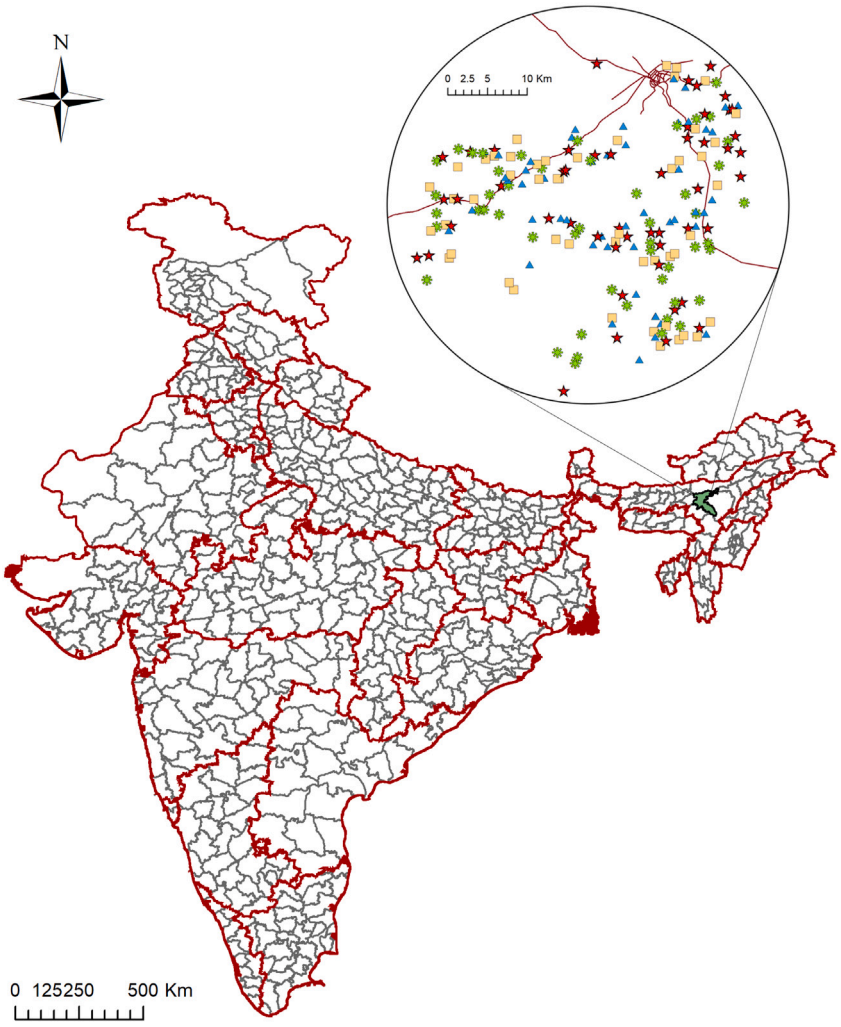


Fig. 1. Study locations.
Notes: The study is set in Nagaon district, within the State of Assam. We drew a sample of 200 villages from the list of villages that Pratham deemed eligible to receive the program. The villages were then randomly assigned to one of the four study arms (50 villages per arm). The map above illustrates the location of Nagaon district within India, as well as an enlarged view of the locations of the 200 study villages. The four different symbols indicate assignment to the different study arms.

Table 1
Comparison of the *Learning Camps* vs Banerjee et al. (2016).

	This study	Banerjee et al. (2016)
Setting	Assam	Uttar Pradesh
Year	2018–2019	2013–2014
Implementer	Pratham	Pratham
Data Collection	J-PAL	J-PAL
Assessment	ASER	ASER
Study length	~18 months	~12 months
LC Duration	32 (30+2) days	50 (40+10) days
Grades covered	1 to 5	3 to 5
Enrollment per school (grades 3-5)	32	72
ASER math (avg)	2.25	2.11
ASER language (avg)	2.84	1.97

Notes: *Enrollment per school* indicates the average number of students enrolled in grades 3 to 5 in the schools included in the sample. *ASER math* and *ASER language* indicate the average level in math and language, respectively, from comparable tests modeled around the ASER test. To ensure comparability, average test scores are calculated using only end-of-school-year assessments (i.e., endline for both studies) for students enrolled in grades 3 to 5 in control-group schools.

Study groups. The *Study Groups* refer to the out-of-school community-managed program. The Study Groups provide a framework for the community to support studying and learning outside of school while keeping costs and time commitments low. The objective was to help students consolidate what they learn in class, leveraging the power of peer support. The study groups were managed by the local community. Pratham introduced the program in the target villages at the start of the school term and mobilized community volunteers, who were typically local mothers. Each volunteer was then assigned a study group, comprising approximately six children of primary school age living in the same neighborhood ("hamlet"). The number of groups in each village thus varied based on the number of children and available volunteers. The structure of the study groups was flexible, with each volunteer determining the frequency and duration of the meetings. During the study-group meetings, the students work with a peer-to-peer learning approach to help each other with the training material. Following the initial mobilization and setup, Pratham conducted only brief monthly visits to the villages to distribute teaching and learning materials for use as group activities. The study groups were also encouraged to engage in additional reading and studying activities, such as completing school homework.[5,6]

[5] Although study groups did not maintain any formal records of their activities, at endline we conducted a qualitative survey with 40 volunteers

Full program. The full program includes both the Learning Camps, implemented in school, and the Study Groups, implemented in the village. The two programs were implemented following the same protocols described above. The combined program aims to foster students' learning in and out of school and to strengthen Pratham's investment in the TaRL approach by actively involving the community in the learning process.

3. Study design

3.1. Sampling and experiment design

The study sample consists of 200 villages randomly selected from a list of eligible villages identified by Pratham in Nagaon district.[7] Each village hosts one public primary school.

Fig. 1 shows the spatial distribution of the study villages, while Table A.2 compares the primary public schools in our sample to the full set of schools across Nagaon district as well as Assam state, using data from the 2017–18 Unified District Information System for Education (UDISE) database, provided by the Ministry of Education. By construction, all schools in our sample provide instruction in Assamese, as that was the language spoken by Pratham staff and the language in which the test was administered, and have a school management committee, as that was one of the criteria considered by Pratham when assessing the potential for community mobilization. Schools included in our sample also tend to have slightly fewer students enrolled across (slightly) more classrooms and to be open for more days per year. There are instead no differences in terms of teacher–student ratio, school funding, availability of electricity, access to playgrounds, availabity of student records, and adoption of the Continuous and Comprehensive Evaluation (CCE) assessment process.[8]

The 200 villages were randomly divided into four groups of equal size: *Learning Camps & Study Groups*; *Learning Camps*; *Study Groups*; *Control*. A second randomization was conducted to determine the timing of the implementation of the Learning Camps, since Pratham could not run the camps at the same time across all villages, due to resource constraints. Out of the 100 villages assigned to receive the Learning Camps (either alone or in combination with the Study Groups), we randomly selected (stratifying by group) half of them to receive the program immediately after baseline data collection (*phase I*), while the remaining half received it approximately 5 months later (*phase II*). Since the endline survey took place at the same time across the whole sample, this introduced random variation in the time elapsed since children's exposure to the learning camps by endline. Table 2 summarizes the experimental design.

Table 2
Experimental design.

			Study groups	
			No	*Yes*
Learning Camps	*No*		50 villages	50 villages
	Yes	*Phase I*	25 villages	25 villages
		Phase II	25 villages	25 villages

Notes: The 200 villages were randomly assigned to one of the four main study arms as shown in the table. Villages assigned to receive the Learning Camps were further randomized into two groups to determine whether they were to host the Learning Camps in the first or second phase of the implementation (see Fig. 2 for the timeline).

3.2. Data

The study uses original survey data collected independently by the research team and administrative data from Pratham. The administrative data, available only for the treatment arms and largely related to the implementation of Learning Camps, are used to assess protocol adherence, measure student participation, and track learning progress based on Pratham's internal assessments.

The variables used in the main analysis were instead collected by an independent research team through three surveys: a school survey, a child survey, and a caregiver survey. The school survey was administered to all 200 head teachers and included a mix of direct questions and observational recordings. The survey collected information on teachers' and students' attendance, school facilities, management and pedagogical practices, and stakeholder interactions. The child survey represents our key source of information and was administered to a random sample of 8 students in each of grades 1 to 4 at baseline (thus expected to be in grades 2–5 at endline),[9] for a total of 32 students per school.[10] Student selection was based on official school enrollment registries, and sampled students who were absent on the survey day were surveyed at home. The child survey included two parts: a learning test module and a short survey. The test module included two tests, each with a mathematics and a language section. The first test (ASER test) mirrored the standard test conducted annually by the ASER Center across India for children aged 5 to 16. ASER-style tests have been extensively used in the literature to measure learning and assess program impact (e.g. Banerjee et al., 2007, 2016). The second test (Test A) was created by the research team based on

across the study villages. The survey revealed significant heterogeneity, but most groups met 2 or 3 times per week for about 2 h, with children primarily focusing on the material shared by Pratham.

[6] The study groups share conceptual features with the volunteer-based program studied in Banerjee et al. (2010), in that both aim to support learning outside of school by mobilizing community members and providing ongoing support from Pratham. However, the two programs differ in key dimensions. In the program studied by Banerjee et al. (2010), volunteers were trained to act as instructors and deliver teaching sessions following a specific pedagogical approach, whereas in our program, the volunteers only serve as organizers of the study groups, and are not expected to deliver teaching sessions, as the groups are designed to stimulate peer-to-peer learning, guided by printed materials provided by Pratham.

[7] Pratham's standard preliminary assessment to identify villages eligible for their programs considers the following three factors: official enrollment levels in the local primary school, primary teaching language, village accessibility, and a qualitative assessment of community mobilization potential (which includes the presence of a school management committee).

[8] Guariso and Björkman Nyqvist (2023) also compare average household characteristics from the national NFHS-V household survey to households included in the study sample. The analysis should be interpreted with caution, as data were collected at different points in time and there is limited overlap among the variables, as NFHS primarily focuses on health. Nevertheless, the comparison shows that the study sample includes relatively more Hindu and scheduled castes households, that households tend to be slightly larger (although with fewer young children), and slightly less educated when considering secondary education (there is no difference when focusing on primary education), compared to the average household in Assam. There is instead no significant difference in terms of wealth, as measured through an index that combines 9 comparable items.

[9] 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 study setting.

[10] If enrollment 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 enrollment 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. Actual enrollment figures turned out to be lower than official figures reported in the central administrative records, and we therefore ended up surveying on average 7 children per grade or 29 children per school.

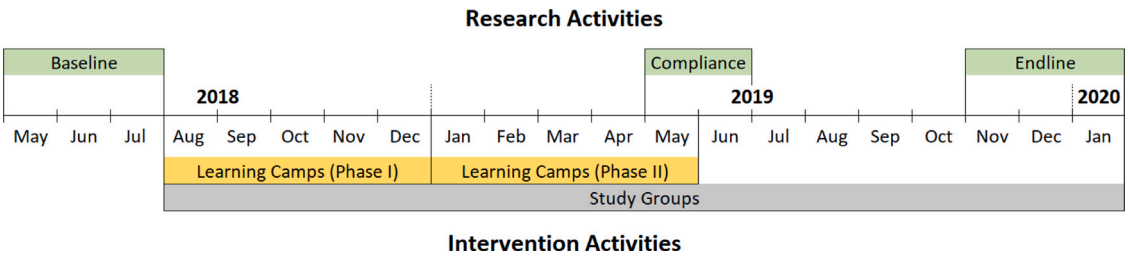


Fig. 2. Timeline.

questions used in other studies in India (Muralidharan and Singh, 2020) and served to enrich the standard ASER assessment. Both ASER and Test A were conducted orally and individually with each sampled child by trained enumerators (see Appendix C for details). However, Test A proved to be relatively easy for the students in the sample and we faced issues of ceiling effects at endline (especially in language).[11] In our analysis, we therefore focus on the standard ASER test as our primary learning measure while also reporting the main results on Test A as well as on a comprehensive measure that takes into account all questions from both tests.[12] The ASER test has the advantage of being directly comparable to national tests as well as to previous TaRL studies in the literature. One concern is that Pratham's Learning Camps, which divide students based on the results of an initial ASER-style test, might be particularly effective in teaching to that test. However, the lack of impact in the study arm exposed to Learning Camps only alleviates this concern in our context. In any case, our results are confirmed when we consider all survey questions, although they become weaker for language.[13]

The students' survey component primarily contained questions about studying and learning practices. At endline, we also included a psychological wellbeing module based on the Child and Adolescent Social and Personal Assessment of Wellbeing (for more details, see Guariso and Björkman Nyqvist, 2023). Finally, at baseline, we selected a random subset of 6 children per grade (totaling 24 children per school) from within our study sample and tracked and interviewed their primary caregivers.[14] The caregiver survey collected information on household characteristics, educational inputs, and educational practices.

3.3. Timeline and study integrity

The baseline survey took place between May and August 2018. Our baseline data show that schools in our sample had, on average, 53 students enrolled, 2 classrooms, and 4 teaching staff (Table 3). Attendance on survey day was on average 70%. About one in four students (25%) reported never studying after school. In terms of learning levels at baseline, the average student in our sample scored between level 1 (beginner) and level 2 (corresponding to 1-digit number recognition in math and letter recognition in language) in the ASER test components. While there is a progression in learning across grades (Figures A.1 and A.2 in the appendix), the overall learning levels are low: about 35% of students in grade 4 were unable to solve subtraction problems, and 66% were unable to read a story, which are the expected standards for a grade 2 student. Table 3 shows that observable characteristics of schools, children, and primary caregivers were balanced across study arms at baseline.[15]

Implementation of the programs started immediately after the collection of baseline data. Study Groups activities started in all targeted villages at the same time and continued throughout the study period while Learning Camps were rolled out in two phases, as described above. A short compliance survey was conducted with head teachers in May and June 2019. The endline survey was conducted approximately 17 months after the program's initiation, between November 2019 and January 2020. At endline we were able to track and survey 93% of the baseline sample (Fig. 2 visually summarizes the study timeline). Attrition rates did not differ across study arms and there was no difference across treatment arms in characteristics of children lost at endline (Table A.3).[16]

We check program implementation and compliance in three ways. First, we matched our sample of schools and children with Pratham's detailed administrative records to confirm that Learning Camps and Study Groups occurred in the assigned villages. Table 4 summarizes the results and shows high adherence to the study design, with the exception of two villages with similar names, originally assigned to the control and the full program intervention, which were mistakenly swapped by Pratham. We adopt a conservative approach in the analysis and rely on the original random assignment, but all results remain consistent when we use the actual assignment.[17] Virtually all Learning Camp days took place as expected (with an average of 28 days out of the expected 30). During the camps, there was an average of 1.6–1.7 volunteers per camp supporting Pratham's staff: one-third of the volunteers were school teachers, while two-thirds were community members. Looking at student-level data in panel B, we see that students, on average, attended about 20 camp days, i.e. about two-thirds of the total expected number of days. When it comes to the study groups, they were created in all target villages, as per study design, with an average

[11] Figures C.5 and C.6 report the distribution of correct answers in mathematics and language, respectively.

[12] We build this measure using a combination of two-parameter logistic (2PL) and three-parameter logistic (3PL) item response theory (IRT) models, following Jacob and Rothstein (2016). Figures C.7 to C.10 show the empirical fit to the estimated Item Characteristic Curve (ICC) at endline. Overall, the graphs show that questions were able to discriminate between students with different achievement levels.

[13] As the next section will show, treatment effects are also confirmed when using only Test A mathematics scores, while they become smaller and non-significant when considering Test A language scores. As mentioned above and shown in Appendix C, Test A language scores suffered from ceiling issues, with half of the students correctly answering 90% of the questions at endline.

[14] The respondent was typically the mother (85%), followed by the father (5%), and the grandmother (3%). Because some schools were smaller than the target size, we surveyed an average of 23 primary caregivers in each village (see Table A.3 in the appendix).

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

[16] At endline one head teacher from the Study Groups treatment arm refused to answer the endline questionnaire (while still giving permission to survey children in the school) and we therefore lack one school survey. Caregivers' attrition mirrors children's attrition (93%).

[17] Table A.4 reports the impact on test scores using the actual assignment.

Table 3
Pre-intervention characteristics of schools, children, and caregivers.

	C	LC & SG	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	7.69	7.64	7.61	7.61	0.77
	(1.59)	(1.55)	(1.53)	(1.56)	
Girl	0.51	0.49	0.50	0.50	0.80
	(0.50)	(0.50)	(0.50)	(0.50)	
Present in school	0.68	0.75	0.74	0.68	0.35
	(0.47)	(0.44)	(0.44)	(0.47)	
Likes going to school [1–5]	4.51	4.54	4.57	4.59	0.33
	(0.89)	(0.85)	(0.81)	(0.78)	
Study outside school	0.71	0.75	0.74	0.76	0.25
	(0.45)	(0.43)	(0.44)	(0.43)	
ASER score (language)	1.75	1.72	1.66	1.63	0.78
	(1.44)	(1.43)	(1.42)	(1.41)	
Reached Grade II level (language)	0.17	0.17	0.16	0.15	0.89
	(0.38)	(0.37)	(0.36)	(0.36)	
ASER score (math)	1.42	1.43	1.40	1.35	0.67
	(0.84)	(0.85)	(0.89)	(0.85)	
Reached Grade II level (math)	0.12	0.12	0.12	0.11	0.89
	(0.32)	(0.33)	(0.33)	(0.31)	
Tracked at endline (share)	0.94	0.93	0.92	0.93	0.59
	(0.24)	(0.25)	(0.28)	(0.25)	
Panel C: Primary Caregivers (N = 4592)					
# HH members	5.24	5.15	5.21	5.30	0.55
	(1.76)	(1.65)	(1.70)	(1.82)	
# children enrolled	1.44	1.42	1.47	1.43	0.61
	(0.82)	(0.78)	(0.82)	(0.84)	
Asset Index	−0.10	0.19	−0.13	0.04	0.23
	(1.89)	(1.78)	(1.81)	(1.83)	
Primary caregiver is literate	0.48	0.53	0.51	0.50	0.73
	(0.50)	(0.50)	(0.50)	(0.50)	
Would like child to go to university	0.44	0.45	0.45	0.49	0.61
	(0.50)	(0.50)	(0.50)	(0.50)	
Pays tuitions	0.19	0.24	0.19	0.21	0.27
	(0.39)	(0.43)	(0.40)	(0.41)	
Tracked at endline (share)	0.95	0.93	0.93	0.94	0.38
	(0.22)	(0.25)	(0.26)	(0.24)	

Notes: Standard deviations shown in parentheses. C, LC & SG, SG, and LC indicate that the observation belongs to one of the following four study groups: control, full program (which included both Learning Camps and Study Groups), Study Groups alone, or Learning Camps alone. The last column shows the *p*-value from testing whether the means are equal across all treatment groups (H_0 := means are equal across groups).

Table 4
Implementation statistics based on Pratham's administrative data.

	LC	LC&SG	SG	p-value ($LC = LC\&SG$)	p-value ($SG = LC\&SG$)
Panel A: School-Level Statistics					
Schools (N)	50	50	50	×	×
Learning Camp (Y/N)	100%	98%	×	×	×
Avg. Camp Days	28.22	27.86	×	0.605	×
Avg. Volunteers	1.71	1.64	×	0.793	×
Teacher Share	38%	36%	×	0.862	×
Study Group (Y/N)	×	100%	100%	×	×
Avg. Study Groups	×	4.92	4.18	×	0.028
Panel B: Student-Level Statistics					
Students (N)	2248	2276	×	×	×
Days attended - Total	20.05	19.83	×	0.149	×

Notes: Data come from Pratham's administrative records. The p-values in the last two columns test the null hypothesis of equal means across intervention arms.

Table 5
Self-reported program exposure.

Panel A: Children	Pratham		Learning Camps			Study Groups	
	Heard of	Interacted	Tested	Teachers	in groups	in village	participated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LC & SG	0.16***	0.36***	0.13***	0.10***	0.35***	0.24***	0.21***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
SG	0.05**	0.21***	0.03**	0.01	0.16***	0.16***	0.14***
	(0.02)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)
LC	0.10***	0.27***	0.08***	0.06***	0.28***	0.01	0.01
	(0.02)	(0.03)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
Mean Control group	0.08	0.34	0.05	0.03	0.24	0.07	0.03
Observations	5328	5328	5328	5328	5328	5328	5328
p-val(LC & SG = SG)	0.00	0.00	0.00	0.00	0.00	0.05	0.05
p-val(LC & SG = LC)	0.02	0.01	0.02	0.04	0.04	0.00	0.00
Panel B: Caregivers	Pratham			Learning Camps			Study
	Heard of	Interacted	Material	In school	TL activities	Diff. gr.	Groups
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LC & SG	0.15***	0.09***	0.36***	0.12***	0.10***	0.23***	0.26***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)
SG	0.09***	0.06***	0.25***	0.08***	0.06***	0.14***	0.23***
	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)	(0.03)	(0.03)
LC	0.08***	0.03**	0.24***	0.05***	0.04***	0.13***	0.04***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
Mean Control group	0.07	0.02	0.10	0.02	0.02	0.05	0.02
Observations	4251	4224	4265	4251	4251	4265	4265
p-val(LC & SG = SG)	0.02	0.19	0.00	0.04	0.03	0.02	0.57
p-val(LC & SG = LC)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: Schools	Pratham		Cearning Camps		Study		
	Heard of	Interacted	Direct	activities	Groups		
	(1)	(2)	(3)	(4)	(5)		
LC & SG	0.58***	0.92***	0.22***	0.62***	0.74***		
	(0.07)	(0.04)	(0.06)	(0.08)	(0.06)		
SG	0.46***	0.78***	0.10**	0.23**	0.55***		
	(0.09)	(0.06)	(0.05)	(0.09)	(0.07)		
LC	0.54***	0.88***	0.22***	0.58***	0.34***		
	(0.08)	(0.05)	(0.06)	(0.08)	(0.07)		
Mean Control group	0.38	0.02	0.02	0.16	0.02		
Observations	199	199	199	199	199		
p-val(LC & SG = SG)	0.04	0.03	0.13	0.00	0.05		
p-val(LC & SG = LC)	0.40	0.47	1.00	0.64	0.00		

Notes: Panel A focuses on the children's survey and the dependent variables are indicators capturing whether the child: ever heard of Pratham (1); ever interacted with Pratham (2); Pratham performed any test (3); Pratham staff replaced the regular teacher (4); students were divided into groups, based on learning level, and combined with students from other grades (5); Study Groups are arranged in the village after school (6); child participated in these groups (7). Panel B focuses on the caregivers' survey and the dependent variables are indicators capturing whether the caregiver: ever heard of Pratham (1); ever interacted with Pratham (2); ever saw Pratham's teaching-learning material (3); Pratham worked in the school (4); Pratham placed the child in a grade different from their usual grade (5); the child worked in groups different from their usual class (6); any Study Group was arranged in the village (7). Panel C focuses on the head teachers' survey and the dependent variables are indicators capturing whether the head teacher or school: ever heard of Pratham (1); ever interacted with Pratham (2); hosted Pratham's Learning Camps (3); Pratham arranged students according to their abilities (4); Pratham arranged after-school study groups in the village (5). LC & SG, SG and LC are indicator variables that take the value 1 if the child's village was assigned to the full program, 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 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.

of 4–5 groups per village. The table also provides a direct comparison across treatment arms, showing no differences in the implementation of Learning Camps between the LC and full-program arms. We observe instead a difference in the study groups, with an average of 0.74 extra groups per village in the full-program arm compared to the SG-only arm. This might be a consequence of Pratham's higher level of engagement in these villages: by running two activities that both require volunteers, Pratham could generate greater interest and participation.

As a second step, we use information from our compliance survey, which was collected when the in-school activities were ongoing or had just been completed. This short survey was designed to monitor implementation fidelity and only covered head teachers. Appendix Table A.5 reports the results and confirms high fidelity: 88% of head teachers from schools assigned to Learning Camps recalled Pratham being at the school, and 98% recalled the camp activities (though some did not recall the exact NGO name), compared with 4% in the Control or Study Groups only arms. Awareness of the Study Groups that took place in the villages (not in the schools) is noisier, but in line with the expected pattern.

Finally, at endline we collected retrospective information from head teachers, children, and caregivers on their interactions with Pratham and exposure to the programs. Given that the survey took place 5–12 months after implementation of the in-school intervention, these measures are expected to be affected by more noise and recall error. We therefore use them primarily to assess whether response patterns vary systematically with treatment assignment. Panel A of Table 5

shows that among children, self-reported awareness of Pratham is low (18% across treatment groups), but the distribution aligns with the experimental design: those in Learning Camps villages are more likely to report camp-related school activities, while those in Study Groups villages are more likely to report the presence of study groups (and their own participation). Panel B shows similar patterns in caregivers' responses, with low reported awareness (18%), higher recognition of Pratham learning materials (38%), and reports of activities consistent with the treatment assignment. Finally, panel C reports head teachers' responses: interaction with Pratham is high in Learning Camps schools (about 92%), though direct awareness of the "Learning Camps" label is lower (24%, rising to 76% when asked indirectly). In Study Groups villages, 67% of head teachers report awareness of the groups. Overall, despite the widespread noise in these self-reported measures, the systematic variation in responses with treatment assignment suggests that the program was implemented as intended at the school and village levels.[18]

Taken together, our checks on baseline balance, attrition, and implementation fidelity alleviate concerns about the study's integrity and indicate that any lack of impact is unlikely to be due to major failures in the study design. At the same time, Table 5 shows that awareness of and (self-reported) participation in the program are generally low. If we compare these measures across the different treatment arms (tests are reported in the bottom rows of each panel), the table highlights that respondents tend to have a higher awareness of and exposure to Pratham's activities in the villages where the full program was implemented, as opposed to villages where only one component was implemented. Given the identical protocols and implementation of the interventions across the combined and individual treatment arms (Table 4), this suggests that combined in- and out-of-school mobilization increased the visibility and uptake of the educational interventions.

3.4. Empirical strategy

Our main estimating equation is:

$$Y_{i,v} = \alpha_1 LC\&SG_v + \alpha_2 SG_v + \alpha_3 LC_v + \Omega X_{i,v} + \epsilon_{i,v} \quad (1)$$

where $Y_{i,v}$ is the outcome for child i, who attends school in village v. $LC\&SG$, SG, and LC, are indicator variables that take value one if the village v was assigned to the full program, to the Study Groups alone, or to the Learning Camps alone, respectively. $X_{i,v}$ is a vector of standard individual-level control variables: age, gender, indicators for grade, and baseline value of the outcome variable $Y_{i,v}$. We cluster standard errors at the village level, accounting for the fact that the intervention varies at that level. The coefficient α_1 shows the average impact of Pratham's full program on the outcome Y, with respect to the control group, while α_2 and α_3 capture the average impact of the two individual components whenever they are implemented alone.

By comparing these three coefficients, we can also test whether the effectiveness of the program differs across the three treatment arms. School- and household-level analyses follow a similar specification, with variables defined at the corresponding level.

In order to form a judgment about the impact of the intervention on a family of n related outcomes, and address potential multiple hypothesis testing concerns, in our analysis we follow Kling et al. (2004) and estimate a seemingly unrelated regression system to derive the average standardized treatment effect (ASTE).

Given the limited program participation, the coefficients α_i should be interpreted as measuring the impact of having the programs implemented in child i's own school or village and not the impact of directly participating in them (intention-to-treat). In Section 5 we will discuss limited participation and resort to an instrumental variable approach to estimate the treatment-on-the-treated effects. More specifically, we will estimate the following 2SLS equation:

$$Y_{i,v} = \beta_1 \widehat{LC_{att}\&SG_{att}}_{i,v} + \beta_2 \widehat{SG_{att}}_{i,v} + \beta_3 \widehat{LC_{att}}_{i,v} + \Omega X_{i,v} + \varepsilon_{i,v} \quad (2)$$

where $LC_{att}\&SG_{att}$ takes the value one if student i attended at least one learning camp day in the school and reported attending the Study Groups in the village at least once, SG_{att} takes the value one if the student reported attending the Study Groups but did not attend any Learning Camp day, and LC_{att} takes the value one if the student attended at least one learning camp day but reported never attending the Study Groups.[19] We instrument the three endogenous exposure variables with the random assignment to the different study arms. However, this analysis should be interpreted with caution, as it rests on the strong assumption that the program only affected learning through direct participation. We believe the most likely direction of the potential bias from violating the exclusion restriction is towards underestimating the true effect of direct exposure, as children not directly exposed might benefit from interactions with exposed children, but we cannot fully rule out other confounding factors. We therefore present this analysis as suggestive.

4. Results

4.1. Student learning outcomes

Table 6 shows the impact of the programs on children's learning outcomes in mathematics and language, our primary outcomes of interest. The table reports four different learning measures: ASER test, an indicator for whether the child reached the learning level expected for a child in grade 2, Test A, and a combination of all questions included in ASER and Test A obtained through Item Response Theory. The estimates show no impact on test scores when the two components (Learning Camps and Study Groups) are implemented individually. Not only are the results not statistically significant, but the estimated coefficients are close to zero.

By design, our study aimed to understand whether the combination of the learning camp program with a community-based component could enhance students' learning. The estimates reported in the first row of Table 6 show that, when considering ASER scores, the combination of the two components significantly increased both mathematics and language test scores by, on average, 0.11 SD (column 1) and 0.09 SD (column 5), respectively, compared to the control group.[20] The average effect size of 0.1 SD corresponds to the median effect size across 270 educational programs implemented in low- and middle-income

[18] We are confident in attributing the imperfect pattern observed in Table 5 to recall errors rather than implementation failures for three reasons. First, when comparing Panel C of Table 5 and Table A.5 we clearly see that the same head teachers, when asked the same questions several months apart, often reported different answers. For instance, 98% of head teachers in treated schools recalled the learning camp activities at the time of the compliance survey, compared to 76%–78% at endline. Second, although we do not have any interim data collected from children and caregivers, we can get a sense of the quality of their self-reported endline information by checking the consistency of their replies across respondents who live in the same village and are enrolled in the same school. In villages where at least one child reported attending the learning camps (study groups) at endline, on average, only 43% (20%) of other children from the same school (village) reported the same. Third, even among students that we know from Pratham's administrative records attended all 30 days of camps, only about 60% report attending them at endline. In Section 5 we will moreover formally show that the observed patterns are unlikely to be driven by spillovers or contamination.

[19] Results remain similar for different definitions of exposure to the Learning Camps, with the estimated coefficients becoming larger as the definition of exposure becomes more restrictive (e.g. attending at least 20 days, or at least 90% of camp days).

[20] We follow the literature and normalize test scores using the mean and standard deviation for the control group.

Table 6
Program impact on test scores.

Dep Var:	Mathematics				Language			
	ASER	Grade II	Test A	IRT	ASER	Grade II	Test A	IRT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LC & SG	0.11**	0.06**	0.10**	0.11**	0.09**	0.04*	0.04	0.07*
	(0.05)	(0.03)	(0.05)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)
SG	−0.01	−0.00	0.01	0.01	0.01	0.02	−0.05	−0.01
	(0.05)	(0.03)	(0.05)	(0.05)	(0.04)	(0.02)	(0.04)	(0.04)
LC	0.02	0.02	0.02	0.02	0.02	0.00	0.01	0.00
	(0.05)	(0.03)	(0.05)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)
Mean Control group	0.00	0.30	0.00	0.00	0.00	0.31	0.00	0.00
R-squared	0.415	0.214	0.459	0.547	0.580	0.314	0.530	0.615
Observations	5328	5328	5328	5328	5328	5328	5328	5328
No. of clusters	200	200	200	200	200	200	200	200
p-val(LC & SG = SG)	0.03	0.06	0.12	0.05	0.07	0.40	0.06	0.06
p-val(LC & SG = LC)	0.08	0.15	0.09	0.08	0.10	0.12	0.45	0.09
p-val(LC & SG = LC+SG)	0.18	0.30	0.34	0.26	0.30	0.59	0.20	0.15

Notes: The dependent variables are: children's ASER test score (columns 1 and 5); an indicator for the student reaching at least Grade II level in the ASER test (columns 2 and 6); children's Test A score (columns 3 and 7); children's test scores obtained by considering all test items from both ASER and Test A, through a combination of two-parameter logistic (2PL) and three-parameter logistic (3PL) item response theory (IRT) models to account for the presence of both open and multiple-choice questions (following Jacob and Rothstein, 2016) (columns 4 and 8). Test scores are always normalized using the mean and standard deviation for the control group. LC & SG, SG and LC are indicator variables that take the value 1 if the child's village was assigned to the full program (which included both Learning Camps and Study Groups), to the Study Groups alone, or to the Learning Camps alone, respectively. All regressions include age, gender, baseline value of the dependent variable, and grade fixed effects (coefficients omitted). The p-values at the bottom of the table are from tests 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.

countries, as reported by Evans and Yuan (2022). In our context, it translates into a 20% (13%) increas in the share of primary school children who achieve minimum standards (i.e. grade 2 level) in mathematics (language) (columns 2 and 6). Over the 17-month study period, the average student in the control group improved by 0.53 SD in mathematics and 0.42 SD in language. Students in the full program arm thus experienced additional gains equivalent to 21% of the control group's mathematics progress and 14% of its language progress.

The p-values reported at the bottom of the table confirm that the full program led to a significantly higher impact on children's test scores than either of the two components alone, with the differences being significant at the 10% level.[21] While the impact of the full program is also larger than the sum of the impact of the two components implemented alone, we cannot reject additivity at conventional significance levels (last row of the table). Our results thus suggest that the combined intervention leads to significant improvements in learning, by providing a higher overall dose of instruction and learning exposure, in- and out-of-school.

Table 6 also reports the treatment effects on learning, as measured by Test A (columns 3 and 7) and by the IRT measure that combines all questions from ASER and Test A. The impact on mathematics is confirmed by these alternative assessments, whereas the effect is smaller (and not statistically significant for Test A) for language, the subject in which Test A suffered serious ceiling issues (see Appendix C for details). Table A.6 in the Appendix shows that results are slightly larger, but broadly similar, when the analysis is restricted to students who were in class on survey day, thus excluding students tracked and surveyed at home.

4.2. Heterogeneity

In its review of the literature, McEwan (2015) finds that only one in ten education studies was assessed more than a month after the interventions ended, thus making it impossible to understand how learning outcomes evolve over time post-intervention. By the time of our endline survey, between 5 and 12 months had passed since the learning camps ended at the schools assigned to receive them. This not only allows us to assess the impact of the interventions some time after their conclusion but also, thanks to our study design, enables us to investigate whether such timing matters. As already explained, the camps were indeed implemented in two different phases due to cost and logistical reasons, and assignment to phase I or phase II was randomized, with the two phases starting roughly 5 months apart (Fig. 2). In Table 7, we focus on our preferred measure of ASER test scores, and in columns 1 and 2, we report the results when we augment Eq. (1) with an interaction between the indicators for assignment to a Learning Camps study arm (*LC&SG* or *LC*) and an indicator for being in the first implementation phase.[22] The estimates show that children in schools that hosted the Learning Camps earlier achieved higher endline test scores. The results should be interpreted cautiously, as coefficients are imprecisely estimated, but the point estimates suggest that the full education program led to learning improvements twice as large for students who received the program earlier on.[23] While several interpretations of this pattern are possible, it is consistent with the dynamic complementarity hypothesis of skill formation (Cunha and Heckman, 2007): earlier exposure to the learning camps may have placed children on a steeper learning trajectory, allowing them to accumulate more knowledge over time and widen the gap with the control group by endline.

The results are more pronounced for the full program arm, which is consistent with at least two additional explanations. First, the Study Groups were introduced in all villages simultaneously, at the start of the school year, regardless of Learning Camp phase. This means that in Phase I villages, the out-of-school component was implemented alongside the Learning Camps from the outset, while in Phase II villages, Study Groups had already been running for roughly five months before the camps began in the schools. If the Study Groups are most effective when they reinforce skills introduced during the camps, the observed

[21] Figure A.3 shows the change in learning levels experienced across the different study arms, while Figures C.2 and C.3 report the endline distribution of the different learning levels by grade and study arm.

[22] Table A.7 shows the same analysis when considering the aggregated IRT test scores. Despite a few minor changes in precision, our conclusions remain unaffected by the change.

[23] Figure A.4 visually illustrates this difference.

Table 7
Heterogeneous effects.

Interaction with... :	...Phase I		...Girl		...Grade		...BL Score	
Dep Var:	Math	Lang.	Math	Lang.	Math	Lang.	Math	Lang.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LC & SG ×...	0.08	0.10*	−0.00	−0.09	0.03	0.01	0.00	−0.01
	(0.07)	(0.05)	(0.07)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)
SG ×...			0.02	−0.07	0.01	0.02	0.01	0.04
			(0.06)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)
LC ×...	0.06	0.06	−0.01	−0.03	0.02	0.03	0.03	0.03
	(0.07)	(0.06)	(0.07)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)
LC & SG	0.07	0.04	0.11*	0.13**	0.05	0.07	0.11**	0.09**
	(0.06)	(0.04)	(0.06)	(0.05)	(0.09)	(0.09)	(0.05)	(0.04)
SG	−0.01	0.01	−0.02	0.04	−0.02	−0.05	−0.01	0.01
	(0.05)	(0.04)	(0.06)	(0.05)	(0.08)	(0.10)	(0.05)	(0.04)
LC	−0.01	−0.01	0.03	0.03	−0.03	−0.06	0.03	0.02
	(0.05)	(0.04)	(0.06)	(0.05)	(0.09)	(0.10)	(0.05)	(0.04)
Mean Control group	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00
R-squared	0.416	0.580	0.415	0.580	0.415	0.580	0.415	0.580
Observations	5328	5328	5328	5328	5328	5328	5328	5328
No. of clusters	200	200	200	200	200	200	200	200

Notes: The dependent variables are children's ASER test scores in Mathematics or Language. LC & SG, SG and LC are indicator variables that take the value 1 if the child's village was assigned to the full program (which included both Learning Camps and Study Groups), to the Study Groups alone, or to the Learning Camps alone, respectively. All regressions include age, gender, baseline value of the dependent variable, grade fixed effects, and the variable being interacted with (coefficients omitted). Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

phase difference could also reflect the benefit of contemporaneous rather than sequential rollout of the two components. Second, the five-month gap between phases means that the two groups of villages received their camps in different seasons, which may have affected attendance and engagement through factors unrelated to the program itself, such as seasonal variation in parental availability or in children's competing demands on their time. Such seasonal variation could have limited the additive impact of the out-of-school component for Phase II villages, irrespective of any learning trajectory effects. Given these alternative explanations we limit our interpretation to the observation that earlier exposure to the Learning Camps is associated with higher test scores by endline.

We find no evidence of differential program impact across any other pre-specified dimension: gender, grade, or starting test score (columns 3–8 of Table 7).[24]

4.3. Impact on children, schools, and caregivers

In order to better understand the impact of the programs, we look at an additional (pre-specified) set of child, caregiver, and school outcomes collected through our surveys. Results are reported in Table 8. Panel A focuses on child-related outcomes. The programs had no impact on school attendance (column 1), indicating that the estimated impact on test scores is not caused by children's increased number of days in school. Children in the Study Groups villages report a higher likelihood of studying after school (column 2). We do not find any discernible effect on students' aspirations for schooling achievements (column 3), on children's overall perception of schooling and learning (column 4), or on children's psychological wellbeing (column 5).[25] Column 6 reports the average standardized treatment effect (ASTE) across the five students variables and shows no overall change in these dimensions.

In Panel B, we show that caregivers in the study group villages were more likely to be involved in the organization of the study groups (column 1), but none of the interventions affected the amount of time caregivers spent with children (column 2), or the time they spent helping with homework (column 3). There was also no impact on household spending on education (column 4). These results are consistent with the purpose of these programs, which were designed to place low demands on parents' time and money. In all three treatment arms, we observe a significant increase in caregivers' interactions with the school (column 5).[26] Column 6 reports the ASTE across all five caregiver variables and shows an overall increase in parental engagement in the villages that hosted the programs. This finding is encouraging, as it is notoriously difficult to increase parental engagement in education, even with high-intensity programs (e.g., Di Maro et al., 2020). However, this also indicates that higher parental engagement does not automatically translate into higher child learning, which is again consistent with recent evidence from a more heavy-handed intervention on parental engagement in Mexico (Barrera-Osorio et al., 2020).

Finally, in Panel C, we report school-level outcomes. The results show that the programs had no impact on teachers' attendance (column 1), while we find an overall negative impact on both school-level physical inputs, which include teaching and learning materials and investments in school infrastructure (column 2), and schools' engagement with stakeholders, as captured by the frequency of the meetings with school management committees (column 3) and government officials

[24] We also conducted a machine-learning heterogeneity analysis, following the split-sample framework of Chernozhukov et al. (2025), using baseline child covariates. Across reading and math outcomes and across all three treatment arms, the procedure found little evidence of stable treatment-effect heterogeneity: the best BLP and GATES fit statistics were uniformly small, and the estimated heterogeneity terms were generally close to zero and not statistically significant.

[25] To capture aspirations, we asked until which grade the child would like to study. To measure school and learning perceptions we asked three different sets of questions and combined all answers using principal component analysis: (1) the importance they attach to education; (2) how much they like school, reading, and mathematics; (3) a set of questions about the perception of schooling and learning. To measure psychological wellbeing we adopt the Child and Adolescent Social and Personal Assessment of Wellbeing tool (CAPSAW), focusing on the two core domains that are relevant for our setting – personal and school-related wellbeing – each covered by eight questions that are then combined in an index through principal component analysis (see Guariso and Björkman Nyqvist 2023 for more details).

[26] Interaction is measured through an index that combines five variables using principal component analysis: attendance to parent–teacher meetings, discussing education with teachers, receiving information from the school, discussing education with the school, and checking children's marks.

Table 8
Program impact on children, caregivers, and schools.

Panel A: Children	Attendance	Study outside School	Study up to	School satisf. (PCA)	Psychological wellbeing	ASTE
	(1)	(2)	(3)	(4)	(5)	(6)
LC & SG	0.01	0.07***	−0.06	0.00	−0.06	0.04
	(0.03)	(0.02)	(0.19)	(0.15)	(0.15)	(0.04)
SG	−0.01	0.05**	−0.18	−0.14	−0.02	−0.00
	(0.04)	(0.02)	(0.16)	(0.16)	(0.15)	(0.04)
LC	0.01	0.04	−0.25	0.01	0.04	0.01
	(0.04)	(0.02)	(0.19)	(0.15)	(0.14)	(0.04)
Mean Control group	0.75	0.77	11.26	0.04	−0.70	
Observations	5328	5328	4329	5108	4838	5328
p-val(LC & SG = SG)	0.61	0.45	0.53	0.37	0.77	
p-val(LC & SG = LC)	0.97	0.18	0.35	0.97	0.48	
Panel B: Caregivers	**Involved in SG**	**Time w. child (%)**	**Help with homework**	**Education expenditure**	**School interac. (PCA)**	**ASTE**
	(1)	(2)	(3)	(4)	(5)	(1)-(5)
LC & SG	0.03***	−0.00	0.12	0.06	0.21**	0.13***
	(0.01)	(0.01)	(0.23)	(0.32)	(0.09)	(0.04)
SG	0.02***	−0.01	0.17	0.05	0.19**	0.10**
	(0.01)	(0.01)	(0.29)	(0.34)	(0.09)	(0.04)
LC	0.00	0.00	0.20	−0.16	0.20**	0.06*
	(0.01)	(0.01)	(0.24)	(0.32)	(0.08)	(0.03)
Mean Control group	0.00	0.24	5.29	3.18	−0.15	
Observations	4265	4251	4250	4265	4246	4265
p-val(LC & SG = SG)	0.33	0.78	0.84	0.98	0.84	
p-val(LC & SG = LC)	0.00	0.64	0.68	0.31	0.92	
Panel C: Schools	**Teachers pres. (%)**	**Material invest. (PCA)**	**Interac. with SMC**	**Interac. with PTA**	**Interac. with MoE**	**ASTE**
	(1)	(2)	(3)	(4)	(5)	(1)-(5)
LC & SG	0.01	−0.40*	−0.20**	−0.17	−0.15**	−0.34***
	(0.05)	(0.23)	(0.09)	(0.11)	(0.07)	(0.11)
SG	0.02	−0.47**	−0.04	0.04	−0.11	−0.18*
	(0.05)	(0.23)	(0.08)	(0.10)	(0.08)	(0.10)
LC	0.06	−0.36	−0.22**	−0.12	−0.14*	−0.28**
	(0.04)	(0.24)	(0.09)	(0.10)	(0.07)	(0.11)
Mean Control group	0.79	0.30	0.84	0.56	0.90	
Observations	199	197	199	186	197	199
p-val(LC & SG = SG)	0.76	0.78	0.09	0.05	0.67	
p-val(LC & SG = LC)	0.22	0.82	0.83	0.64	0.89	

Notes: Panel A focuses on the children's survey and the dependent variables are as follows: child was in school on survey day (1); child studies outside school (2); child would like to study until grade X (3); principal component analysis (PCA) across 19 variables on child's school satisfaction (4); psychological wellbeing as captured through CAPSAW assessment (5); ASTE of columns 1-5 (6). Panel B focuses on the caregiver's survey and the dependent variables are as follows: caregiver 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 the child with homework (3); expenditure on education-related items (4); PCA across 5 variables on caregiver's interaction with the school (5); ASTE of columns 1-5 (6). The varying number of observations is due to some respondents not knowing the answer to some specific questions. Panel C focuses on the head teacher's survey and the dependent variables are as follows: share of teachers present at school on survey day (1); PCA over 4 variables capturing physical investments undertaken in the school in the previous 14 months (2); any recent School Management Committee meeting (3); any recent parent–teacher meeting (4); any recent visit from the Ministry of Education (5); ASTE of columns 1-5 (6). All regressions in Panels A and B include age, gender, baseline value of the dependent variable (whenever available), and grade fixed effects (coefficients omitted). LC & SG, SG and LC are indicator variables that take the value 1 if the child's village was assigned to the full program, to the Study Groups alone, or to the Learning Camps alone, respectively. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

(column 5). By contrast, we find no impact on the frequency of meetings between parents and teachers (column 4). Column 6 reports the ASTE across all five school variables and shows an overall reduction in school-level inputs, which appears especially large for schools that hosted the learning camps, suggesting that the program might have crowded out school and government inputs in the education process. The fact that agents, in this case, schools and the government, adjust their investments in response to support programs appears in line with what Das et al. (2013) find in the context of school grants in India and Zambia.[27] To the extent that these inputs support the children's learning process, such a substitution effect might have attenuated the impact of the program. We return to this point in the discussion below.

[27] In that case, the substitution effect was driven by households, reducing their investments in response to anticipated grants distributed to the schools.

4.4. Cost-effectiveness

We use data from Pratham to calculate cost-effectiveness for each intervention arm, estimating the cost of serving 50 villages across the three versions of the program: full program, Learning Camps only, and Study Groups only. We consider all costs associated with Pratham staff salaries, development of the material, training, travel, and any additional costs associated with the rental of office space, utilities, and regular workshops.[28] The analysis reported in Table 9 highlights three facts. First, as expected, the Learning Camps component is significantly

[28] We do not consider the opportunity cost of parental time associated with study groups, because we do not observe in our time use data any significant shift in parental time use. Moreover, the potential cost of having one parent manage the study group could be compensated by the benefit to other parents

Table 9
Cost-effectiveness analysis.

	LC & SG	LC	SG
	(1)	(2)	(3)
Total costs (yearly)	**34,253**	**31,756**	**19,853**
Personnel	23,685	23,685	13,293
TLM	5,369	3,454	3,398
Training	2,619	2,037	582
Travels	614	614	614
Other Costs	1,966	1,966	1,966
# years	1.33	1.33	1.33
# villages served	50	50	50
# children served per village	53	53	53
Avg cost per student	17.2	15.9	10.0
Avg learning gains (SD)	0.09 – 0.11	0.018 – 0.023	−0.01 – 0.01
Cost per 0.1 SD gain	15.0 – 18.3		
Additional SD per 100 US$	0.55 – 0.67		

Notes: All measures are expressed in US$. *TLM* indicates the Teaching and Learning Material developed and printed by Pratham. *Travels* include travel costs for Pratham's supervisors, while the costs for individual staff members to reach the different villages are included in their salaries (under Personnel). *Other costs* include fixed costs such as office space, utilities, and workshops.

more expensive (60% more) than the Study Groups component, which relies on community resources. Second, there are strong cost synergies that can be exploited by combining the two components. The average cost was 15.9 US$ per student for the Learning-Camps-only program, 10 US$ per student for the Study-Groups-only program, and 17.2 US$ per student for the combined program. This means that in the context of an expansion of the Learning Camps program, which we are currently witnessing across India and the African continent, it costs very little to add the Study Group component, which, in turn, can lead to significant learning gains.

Finally, using the learning gains reported in Table 6, we estimate the cost of increasing learning outcomes by 0.1 SD for the full educational program to range between 15 US$ and 18.3 US$ per student. This implies an impact ranging from 0.55 SD to 0.67 SD per 100 US$ invested (under the simplifying assumption of linear returns). To put these numbers in perspective, the program falls right in the middle of the distribution of 27 education programs aimed at improving learning outcomes and evaluated using randomized controlled trials, for which J-PAL has assessed cost-effectiveness (Bhula et al., 2013).

5. Discussion

Our analysis finds positive average effects on learning when the two program components are implemented jointly, but no detectable effects when they are delivered in isolation. In particular, the null effect of the Learning Camps alone contrasts with some of the prior evidence in the literature. In this section, we examine potential explanations for these results, including implementation quality, spillovers and contamination, take-up and effective exposure, school-level behavioral responses, and contextual factors.

Implementation quality (Learning Camps)

As discussed in Section 3.3, protocol adherence was nearly perfect, except for a mix-up between two schools (one full program and one control school) with similar names. Administrative records and the evidence reported in Table 4 indicate that implementation closely followed the experimental design, with similar delivery of Learning Camps across arms and slightly more study groups formed in the full program arm, likely reflecting higher engagement when both components were combined.

However, while high protocol adherence ensures internal validity, it does not necessarily guarantee high implementation quality. We do not have direct measures of implementation quality, as we deliberately avoided introducing additional monitoring systems to preserve standard implementation practices and minimize external influence. But for the Learning Camps, we can draw on Pratham's administrative data to provide indirect evidence on implementation quality. As part of the program, children are assessed by Pratham at the beginning of the first camp (to group them by level) and at the end of each 10-day camp (to track learning progress). Figure A.5 in the Appendix reports mean gains by grade in mathematics and language, standardized relative to the baseline camp assessment.[29] Across the three camps (spanning roughly four months), Pratham's administrative data show average gains in the ASER test of 1.9 SD in mathematics and 0.9 SD in language. These magnitudes are substantially larger than those observed in our independent endline assessment: control-group students improved by 0.61 SD in mathematics and 0.46 SD in language over the full 17-month study period (with respect to the baseline levels). The difference in learning gains between Pratham's learning camps data and our endline data is partly to be expected. Pratham's tests are administered immediately after each camp cycle, focus directly on skills practised during the camps, and are repeated within a short time frame. By contrast, our endline test took place at least 5 months after the last camp. Despite these caveats, the administrative data suggest relevant short-term learning progress during camp implementation. Importantly, gains are broadly similar across both Learning Camp treatment arms (*LC* versus *LC&SG*) and across implementation phases, with the only statistically significant difference being slightly larger language gains in Phase I ($p < 0.05$).

Figure A.6 in the Appendix uses additional administrative data from Pratham to compare mathematics gains across other locations (not part of our RCT) in the states of Assam, Rajasthan, and Uttar Pradesh, where the Learning Camps were implemented during the same period. Learning gains in our study sample are clearly comparable to those observed in other locations, except for slightly slower progress at the highest (division) level. This is consistent with Pratham's internal reports, which did not identify unusual operational challenges during implementation.

Overall, the evidence presented here suggests that weak implementation quality is unlikely to explain the absence of effects of the Learning Camps in the single-component arm.

Spillovers and contamination

We next examine whether cross-village spillovers may have attenuated differences between treatment and control arms. Using the geographic

who do not need to look after their children and can use their time to do other productive activities.

[29] Grade 1 students are tested only at the start of camp 1 and at the end of camp 3.

coordinates of the schools, we test whether self-reported program exposure and endline test scores vary with distance to the nearest village assigned to a different treatment status (distinguishing between Learning Camps and Study Groups). The results, reported in Table A.8, show no evidence that students located closer to villages with a different treatment assignment were more likely to report exposure to that program, nor that their endline performance differed systematically. These results suggest that cross-village spillovers or contamination are unlikely to play a role in our results.

Participation constraints

Limited participation and engagement are common challenges across education programs. Increasing participation was a key reason that led Pratham to develop its Learning Camps model to be implemented during regular school hours rather than after school (Banerjee et al., 2010). Nevertheless, students might still miss school and camp days. We compute exposure to the Learning Camps using Pratham's detailed attendance sheets. We find that most students (93%) assigned to the Learning Camps attended at least one camp day at school.[30] On average, students attended 64% of the camp days and the distribution was similar across Learning Camps only and the joint program schools, as shown in Figure A.7 in the appendix.

The Study Groups took place in the village, after school hours, and children's participation was voluntary. Participation was therefore expected to be lower than in the Learning Camps. The groups were administered by volunteers who did not keep any formal administrative records. We therefore rely on self-reported endline data, which indicate that the share of students who participated in a study group at least once is 17% in the Study-Groups-only arm, 24% in the full program arm ($p = 0.05$ for the difference), and 4% in the Learning Camps and control arms. While greater engagement with Study Groups may have contributed to higher learning gains, the difference is relatively small, and the evidence appears more consistent with the combined intervention increasing overall exposure to learning activities by combining participation across the two components, rather than by substantially raising participation within each component.

In all cases, partial exposure is likely to attenuate the intent-to-treat effects estimated above. In our dataset, program exposure is indeed strongly correlated with endline performance. For instance, students in the top 10% of the Learning Camp attendance distribution (who attended 28 days, on average), by endline, performed 0.38 SD (0.12 SD) better in mathematics (language) than those in the bottom 10% of the distribution (who attended 3 days, on average), within the Learning Camp only arm, after controlling for baseline test scores and standard covariates. Despite the battery of controls, however, these associations likely reflect, at least in part, selection into program participation. To address this selection, we implement the instrumental variables strategy described in Section 3.4, using treatment assignment as an instrument for program participation. The estimates reported in Table 10 indicate large effects for students in the combined arm that were exposed to both programs, with learning gains of 0.48 and 0.38 SD in mathematics and language, respectively (based on ASER scores). These effects are well within the range of treatment-on-the-treated effects observed in previous studies of targeted interventions (Angrist and Meager, 2023).[31] The effects in the individual *LC* and *SG* arms instead remain close to zero. These results thus suggest that limited participation attenuates the average effects observed in the main analysis but cannot explain the absence of detectable impacts in the *LC* (or *SG*) arm.

School responses

Table 8 showed that schools exposed to the program – particularly those hosting Learning Camps – reduced some regular school inputs, including infrastructure investments and interactions with stakeholders. While such substitution could attenuate treatment effects, it is unlikely to have a dramatic impact on our results. Existing evidence suggests that small changes in standard school inputs tend to have limited effects on students' learning outcomes (e.g., Das et al., 2013; Glewwe and Muralidharan, 2016). Consistent with this, at baseline, we find no significant association between these inputs and learning outcomes across our sample.

Context-specific factors

Finally, we consider the possibility that specific contextual features might have limited the effectiveness of the interventions. As described above, Banerjee et al. (2016) evaluate a closely related TaRL model in Uttar Pradesh in 2013, finding learning gains of 0.7 SD. To understand whether the different results might be explained by observable differences across the two samples, we combine our data with those of Banerjee et al. (2016).[32] Table A.9 shows that the difference in impact cannot be directly attributed to differences in initial test scores, age, grade, gender, time between surveys, school size, classroom-average test scores, or the variance of classroom test scores (the regression analysis reported in the table controls for these dimensions as well as for their interactions with the program indicator).

Although our study shares many key features with Banerjee et al. including the pedagogical model, implementing partner, data collection agency, and assessment tool, there remain important differences, both in terms of program and context. As explained above (Table 1), the number of camp days was substantially reduced from 50 to 32. This allows for lower costs and facilitates scaling up the intervention, but also reduces exposure to program learning. In terms of context, Assamese schools are, on average, less than half the size of the Uttar Pradesh sample, and the average starting level in language is almost one point higher (on the ASER five-point scale). Although the empirical analysis above does not show clear evidence that program effectiveness was influenced by these dimensions, they point to substantial differences across the two samples, which might have influenced program effectiveness in other unobservable ways. Also the broader Indian education context has substantially changed over the past decade. At the time of the Banerjee et al. study, the Learning Camps represented a marked departure from prevailing classroom practices. Since then, Pratham has partnered with state governments across India, leading to a general dissemination of the TaRL principles and methods aimed at integrating the model into public education systems (Banerjee et al., 2016; Banerji and Mukerji, 2017; Banerji et al., 2020). Consistent with this diffusion, 92% of control-group head teachers in our setting report sometimes combining school grades and grouping students by learning level. The Learning Camp in our treatment schools might thus have represented a smaller change in pedagogical method. Having said this, we do not believe this can fully explain the lack of impact in the LC-only arm. Although our survey data document awareness among head teachers of some of the TaRL principles, we have no direct evidence of any actual application of those principles. Earlier experimental evidence also shows that when the TaRL approach is implemented in schools by regular teachers, it has only a limited impact (if any) on students' learning, whereas substantial gains emerge when it is implemented by NGO staff (Banerjee et al., 2016). Consistent with this, baseline learning levels were generally low across our sample and in line with test scores in other settings (see Figure A.6), suggesting that pre-existing regular instruction was not particularly effective in improving student outcomes. This makes it unlikely that any potential pre-existing

[30] Although we cannot fully rule out omissions and entry errors, we manually refined the matching between our data and the administrative records to ensure maximum flexibility. The matching took into account information on school, grade, age, gender, first and second names. Overall, we successfully matched 2567 out of 2760 children included in our sample, with similar matching success across the Learning-Camps-only and joint-program (LG&SG) study arms (p-value = 0.717).

[31] In line with the results of Table 6, the effects become weaker when considering Test A results in Language.

[32] We thank the authors for making their dataset available to us.

Table 10
Program impact on test scores — 2SLS.

Dep Var:	Mathematics				Language			
	ASER	Grade II	Test A	IRT	ASER	Grade II	Test A	IRT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LC & SG	0.48**	0.23*	0.45*	0.45**	0.38*	0.20*	0.19	0.35*
	(0.24)	(0.13)	(0.24)	(0.22)	(0.20)	(0.12)	(0.21)	(0.20)
SG	−0.06	−0.01	0.09	0.04	0.04	0.14	−0.34	−0.11
	(0.38)	(0.21)	(0.39)	(0.35)	(0.32)	(0.17)	(0.34)	(0.31)
LC	0.00	0.01	0.00	0.01	0.00	−0.00	−0.01	−0.02
	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)	(0.03)	(0.05)	(0.05)
Mean Control group	0.00	0.30	0.00	0.00	0.00	0.31	0.00	0.00
R-squared	0.407	0.210	0.456	0.543	0.576	0.314	0.519	0.606
Observations	5328	5328	5328	5328	5328	5328	5328	5328
No. of clusters	200	200	200	200	200	200	200	200
p-val(LC & SG = SG)	0.21	0.32	0.42	0.31	0.36	0.78	0.18	0.20
p-val(LC & SG = LC)	0.08	0.13	0.10	0.08	0.11	0.14	0.41	0.11
p-val(LC & SG = LC+SG)	0.26	0.39	0.47	0.37	0.41	0.80	0.21	0.22
F-stat. LC & SG	35.53	35.49	35.61	35.62	35.83	36.18	35.72	35.93
F-stat. SG	22.07	22.64	21.80	21.72	22.18	22.63	22.11	22.02
F-stat. LC	1598.51	1574.24	1588.19	1593.29	1602.67	1626.61	1602.53	1617.86

Notes: The structure of the table mirrors Table 6 in the main text. In this case, however LC & SG, SG and LC capture direct participation in the program, measured as follows: *LC & SG* takes the value 1 if the child attended at least one day of LC AND reported ever attending SG; *SG* takes the value 1 if the child reported ever attending SG BUT did not attend any day of LC; *LC* takes the value 1 if the child attended at least one day of LC BUT reported never attending SG. These variables are instrumented using assignment indicators for the different treatment arms. All regressions include age, gender, baseline value of the dependent variable, and grade fixed effects (coefficients omitted). F-stats from the first stage are reported at the bottom of the table. Standard errors clustered by village in parentheses. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

integration of TaRL principles in the regular curricula of the schools could make the Pratham-led intervention redundant. Nevertheless, the combination of programmatic and contextual differences between this study and that of Banerjee et al. might help explain why the Learning Camp intervention in our study was insufficient to generate measurable improvements in children's learning, underscoring the importance of continuous monitoring and adaptation, even for well-established interventions.

Overall, the evidence provided in this section narrows the set of plausible explanations for the absence of impacts when the two components are implemented in isolation. We find little support for implementation failures, contamination, or spillovers as primary drivers. Instead, the results are consistent with participation constraints, school-level behavioral responses, and context-specific limits jointly contributing to the limited effectiveness of standalone interventions. When combined, however, the two components increase overall exposure to learning activities, in- and out-of-school, leading to significant gains.

6. Conclusion

Targeted pedagogical interventions have emerged as one of the most widely used approaches to improve learning outcomes in low-income settings (Global Education Evidence Advisory Panel, 2020). While existing evidence shows that such interventions can be highly effective, their impacts vary substantially across studies and contexts (e.g., Angrist and Meager, 2023). Understanding how to sustain and adapt these programs across settings therefore remains an important open question. In this paper, we examine whether combining a standard in-school targeted instruction program with a community-managed out-of-school learning intervention can enhance effectiveness and generate larger learning gains.

Our results show that, in our setting, neither the in-school component nor the community-based study groups generate detectable improvements in learning when implemented independently. However, when the two components are delivered jointly, the program increases children's test scores in mathematics and language by about 0.1 standard deviations. The absence of effects in the standalone arms appears to reflect a combination of participation constraints, school-level behavioral responses, and context-specific features. Bundling the two components increases exposure to learning activities in- and out-of-school, leading to significant learning gains.

These results speak to the multi-dimensionality of the learning process and show that educational programs that intervene on several dimensions at the same time can lead to significant learning gains. They also point to the potential of out-of-school, community-based activities to support and extend in-school learning, despite having received relatively little attention both in policy and in the literature. Importantly, this can be achieved at relatively low cost: thanks to cost synergies, the combined program appears competitive with other effective education interventions. In line with our results, the implementing organization, Pratham, recently included the out-of-school component in its flagship educational model across India.

Future research should continue exploring how educational programs can be adapted and strengthened across different contexts. Repeated experimentation has allowed approaches like TaRL to be progressively refined and expanded over time (e.g. Banerjee et al., 2016). Our findings highlight that even well-established pedagogical approaches may need adjustments when implemented in new contexts. A recent study by Duflo et al. (2024) takes a step in this direction, by evaluating alternative TaRL models in Ghana and finding modest but positive impacts ranging from 0.08 to 0.15 standard deviations. Our results suggest that one way to improve the effectiveness of targeted instruction programs is to increase students' exposure to learning activities, by combining the traditional in-school pedagogical model with out-of-school, community-led, programs. In our context, the community-based activities were supported by regular inputs from the NGO, whereas recent evidence from Mozambique suggests that lighter-touch, community-driven models may be less effective when local capacity is limited (Chimbutane et al., 2026). Increasing students' exposure to learning by combining in-school and out-of-school activities is a promising approach to strengthening educational outcomes, but further research is needed to identify the conditions under which such approaches can be implemented successfully across contexts.

CRediT authorship contribution statement

Martina Björkman Nyqvist: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Andrea Guariso:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors have no conflicts of interest to disclose.

Financial support for this study was provided by: Carl Bennet AB, the Swedish Research Council (2016-05615) and the Mistra Foundation, TCD Arts and Social Sciences Benefactions (ASSB) Fund.

The study was registered with the AEA RCT registry (0002817) and received ethical approval from IFMR Human Subjects Committee (IRB00007107) and TCD Ethics Review Board (05062018).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2026.103906.

Data availability

Data and code for replication are available for download from Mendeley Data at https://doi.org/10.17632/mtf2x29p69.1.

References

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

Angrist, N., Meager, R., 2023. Implementation matters: generalizing treatment effects in education. Blavatnik School of Government Working Paper Series BSG-WP-2023/05. Version 2.

ASER, 2016. Annual status of education report (rural) 2016. Pratham Organization. http://img.asercentre.org/docs/publications/ASER%20Reports/ASER%202016/aser_2016.pdf.

ASER, 2018. Annual status of education report (rural) 2018. Pratham Organization. http://img.asercentre.org/docs/ASER%202018/Release%20Material/aserreport2018.pdf.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., Rubio-Codina, M., 2020b. Estimating the production function for human capital: results from a randomized controlled trial in Colombia. Am. Econ. Rev. 110 (1), 48–85.

Attanasio, O., Meghir, C., Nix, E., 2020a. Human capital development and parental investment in India. Rev. Econ. Stud. 87 (6), 2511–2541.

Avvisati, F., Gurgand, M., Guyon, N., Maurin, E., 2014. Get. Parents Involv.: A Field Exp. Deprived Sch. Rev. Econ. Stud. 81 (1), 57–83.

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., Banerji, R., Duflo, E., Glennerster, R., Khemani, S., 2010. Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. Am. Econ. J.: Econ. Policy 2 (1), 1–30.

Banerjee, A., Cole, S., Duflo, E., Linden, L., 2007. Remedying education: evidence from two randomized experiments in india. Q. J. Econ. 122 (3), 1235–1264.

Banerji, R., Agarwal, A., Lakshman, S., 2020. Teaching at the right level: From concern with exclusion to challenges of implementation. Background paper prepared for the 2020 UNESCO Global Education Monitoring Report "Inclusion and education".

Banerji, R., Berry, J., Shotland, M., 2017. The impact of maternal literacy and participation programs: Evidence from a randomized evaluation in India. Am. Econ. J.: Appl. Econ. 9 (4), 303–337.

Banerji, R., Mukerji, S., 2017. Working together to teach at the right level. Article published on Mint on 13 Dec 2017 and available at https://www.livemint.com/Opinion/oZw3aIX99bMUEVV3ycPTPJ/Working-together-to-teach-at-the-right-level.html.

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

Cabezas, V., Cuesta, J.I., Gallego, F.A., 2021. Does short-term school tutoring have medium-term effects? Experimental evidence from Chile. Instituto de Economía, Pontificia Universidad Católica de Chile, Documento de Trabajo No. 565.

Chernozhukov, V., Demirer, M., Duflo, E., Fernández-Val, I., 2025. Fisher–Schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. Econometrica 93, 1121–1164.

Chimbutane, F., Karachiwalla, N., Herrera-Almanza, C., Leight, J., Lauchande, C., 2026. The effect of teacher training and community literacy programming on teacher and student outcomes. J. Dev. Econ. 178, 103578.

Chiplunkar, G., Dhar, D., Nagesh, R., 2023. Can bridging the learning gap improve test scores on a key national exam? Evidence from a remedial education program in Indian secondary schools. Darden Business School Working Paper, No. 3932347. SSRN Working Paper.

Cunha, F., Heckman, J., 2007. The technology of skill formation. Am. Econ. Rev. 97 (2), 31–47.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., Sundararaman, V., 2013. School inputs, household substitution,and test scores. Am. Econ. J.: Appl. Econ. 5 (2), 29–57.

de Barros, A., Fajardo-Gonzalez, J., Glewwe, P., Sankar, A., 2024. The limitations of activity-based instruction to improve the productivity of schooling. Econ. J. 134 (659), 959–984.

Di Maro, V., Leeffers, L., Serra, D., Vicente, P.C., 2020. Mobilizing parents at home and at school: An experiment on primary education in angola. Novafrica Working Paper No. 2002.

Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. Am. Econ. Rev. 101 (5), 1739–1774.

Duflo, A., Kiessel, J., Lucas, A.M., 2024. Experimental evidence on four policies to increase learning at scale. Econ. J. 134 (661), 1985–2008.

Evans, David K., Yuan, Fei, 2022. What we learn about girls' education from interventions that do not focus on girls. World Bank Econ. Rev. 36 (1), 244–267.

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, Elsevier, pp. 653–743.

Global Education Evidence Advisory Panel, 2020. Cost-effective approaches to improve global learning: what does recent evidence tell us are 'smart buys' for improving learning in low- and middle-income countries? The World Bank.

Government of India, 2009. Right of Children to Free and Compulsory Education (RTE) Act. Ministry of Education.

Guariso, A., Björkman Nyqvist, M., 2023. The impact of the COVID-19 pandemic on children's learning and wellbeing: Evidence from India. J. Dev. Econ. 164, 103133.

Jacob, B., Rothstein, J., 2016. The measurement of student ability in modern assessment systems. J. Econ. Perspect. 30 (3), 85–108.

Kling, J.R., Liebman, J.B., Katz, L.F., Sanbonmatsu, L., 2004. Moving to opportunity and tranquility: Neighborhood effects on adult economic self-sufficiency and health from a randomized housing voucher experiment. Harvard Kennedy School Working Paper, No. RWP04-035.

Lai, F., Luo, R., Zhang, L., Huang, X., Rozelle, S., 2015. Does computer-assisted learning improve learning outcomes? Econ. Educ. Rev. 47, 34–48.

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.

Le Nestour, A., Moscoviz, L., Sandefur, J., 2022. The long-run decline of education quality in the developing world. CGD Working Paper, No. 608.

Linden, L.L., 2008. Complement or substitute? the effect of technology on student achievement in india. InfoDev Working Paper, No. 17.

McEwan, P.J., 2015. Improving learning in primary schools. Rev. Educ. Res. 85 (3), 353–394.

Moscoviz, L., Evans, D., 2022. Learning loss and student dropouts during the COVID-19 pandemic: A review of the evidence two years after schools shut down. CGD Working Paper, No 609.

Muralidharan, K., 2017. Field experiments in education. In: Handbook of Field Experiments, vol. 2, pp. 323–385.

Muralidharan, K., Singh, A., 2020. Improving public sector management at scale? Experimental evidence on school governance in India. NBER Working Paper, No. 28129.

Muralidharan, K., Singh, A., Ganimian, A.J., 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. Am. Econ. Rev. 109 (4), 1426–1460.

Romero, M., Chen, L., Magari, N., 2022. Cross-age tutoring: Experimental evidence from Kenya. Econ. Dev. Cult. Chang. 70 (3), 1133–1157.

Singh, A., Romero, M., Muralidharan, K., 2024. COVID-19 learning loss and recovery: Panel data evidence from India. J. Hum. Resour. 590.

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