Teaching at the Right Level: Evidence from Randomized Evaluations in India

Abhijit Banerjee
MIT, NBER, and BREAD

Rukmini Banerji
ASER Centre

James Berry
Cornell University

Esther Duflo
MIT, NBER, and BREAD

Shobhini Mukherji
J-PAL

Marc Shotland
J-PAL

Michael Walton
Harvard University[*†]

September 2014

Abstract

Despite gains in primary school enrollment in India over the past decade, learning levels are stagnating. In this study, we evaluate a set of models to mainstream and scale up the non-governmental organization Pratham's approach to teaching children according to actual learning levels rather than a prescribed syllabus. The methodology has previously been shown to be effective when implemented by community volunteers outside of school. Through three new randomized controlled trials, we evaluate a set of interventions that aim to integrate the methodology into government schools. In two separate evaluations in different states, we find that the Pratham methodology is largely ineffective when implemented without the assistance of volunteers. In a third state, where the methodology was implemented with the support of government monitors, the intervention had significant impacts on reading scores.

[*]banerjee@mit.edu, rukmini.banerji@pratham.org, jimberry@cornell.edu, eduflo@mit.edu, shobhini.mukerji@ifmr.ac.in, shotland@mit.edu, Michael_Walton@hks.harvard.edu

[†]Thanks to Richard McDowell, Harris Eppsteiner, and Madeline Duhon for research assistance; to Anupama Deshpande, Blaise Gonda, John Firth, Adrien Lorenceau, Jonathan Mazumdar, Paribhasha Sharma, Joseph Shields, Zakaria Siddiqui, Yashas Vaidya and Melanie Wasserman for field management; to Harini Kannan and Diva Dhar for supervision; and to Shaher Bhanu Vagh for the educational test design and analysis. Special thanks to Madhav Chavan, Sanjay Kumar Singh and many others in Pratham for their openness and engagement, and to the William and Flora Hewlett Foundation, the International Initiative for Impact Evaluation, the Government of Haryana, and the Regional Centers for Learning on Evaluation and Results for their financial support and commitment.

1 Introduction

During the last decade, enrollment rates at the primary level have rapidly increased in most developing countries. Among 6 to 13 years old in India, for example, only 4.3% were not enrolled in school in 2009 (IMRB, 2010). Unfortunately, while more and more children attend school, learning levels are stagnating. In 2005, the first ASER survey, a nationwide survey on learning assessment found that, among 7-14 year olds, 35 percent could not read a simple paragraph, and 41 percent could not do simple subtraction, competencies that according to government guidelines should be achieved by grade 2 (ASER Centre, 2005). The survey has been conducted every year since and, until 2013, has shown essentially no improvement in students' learning levels, even though enrollment rates have steadily increased over the period. This suggests that schools are teaching next to nothing. Similar findings have come from large surveys in Pakistan, Kenya, Tanzania, Uganda and Ghana (Andrabi et al. 2007, Uwezo 2014, NEA 2008).

Motivated by this lack of progress, a number of experimental studies have shown that changes in pedagogy can lead to significant improvements in learning levels. In particular, interventions which focus on teaching children at the appropriate level, such as remedial education (Banerjee et al. 2007), computer assisted learning (Banerjee et al. 2007) and tracking (Duflo et al. 2011) have systematically found large effects on learning levels, which were often sustained over time. However, with the exceptions of an on-going experimental project conducted by the Government of Ghana (IGC, 2014), and with changes in the ministry of education in the State of Punjab in India, such policies have, by and large, not been adopted by school systems in developing countries. Quite to the contrary, the Right to Education Act, passed in India in 2009, prohibits exams in primary schools, mandates automatic promotion and the completion of the prescribed curriculum in each class, defines the quality of education exclusively in terms of inputs (i.e. teacher salary, physical infrastructure, etc.), and sets standards which could severely restrict informal schools or cheap private schools that cater to children whose parents chose these alternatives. This is quite emblematic of a view of what constitutes a proper education that is typical in many developing countries and leaves very little room for adjusting teaching to a very heterogeneous population.

In this study, we evaluate different models to mainstream and scale up the non-governmental organization Pratham's methodology, which has been implemented through a program called Read India. In this approach, activities include focusing on a specific target group of children, articulating clear learning goals, using appropriate teaching-learning activities and materials, and organizing children into groups by ability level. Ongoing simple measurement of children's ability to read, write, comprehend and do basic arithmetic and solve problems is used to assess progress.

A previous study (Banerjee et al. 2010) showed that, when implemented with carefully monitored but lightly trained volunteers (given only two weeks of training) in Uttar Pradesh, the program delivered impressive results in improving the reading ability of children who attended the program.

In this study, we compare four different models for delivering the Pratham content in schools in the states of Bihar, Uttarakhand, and Haryana. In the first model (implemented in Bihar), teachers were trained to provide remedial education during summer camps (children in need were supposed to be identified in school). In the second model (implemented in Bihar), materials were distributed to schools without any additional support. In the third model (implemented in all three states), teachers were trained to teach with the Pratham method during the school year, and received the appropriate material; in the fourth model, both teachers and volunteers were trained. The fourth model was implemented in Bihar and Uttarakhand, and the volunteers had different roles in each state: in Bihar, the volunteers worked in the community while in Uttarakhand, they were supposed to help teachers at school. Villages or schools were randomly selected to receive one model or the other (or to be a control group), and data was collected on all of them at baseline, after one year, and, in the case of Bihar and Uttarakhand, after two years.

The results confirm the effectiveness of the program when implemented by volunteers in the communities. In this treatment group (implemented only in Bihar and in combination with teacher training), results significantly improved relative to the control group. Overall, learning levels increased by 0.13 standard deviations in Hindi, and 0.10 standard deviations for math. Most importantly, the results show that mainstreaming the program through the regular school system, even with the explicit collaboration of the States' education departments, is difficult without careful

top-down monitoring. First, in Bihar and Uttarakhand, the program which focused on training teachers for implementing the pedagogy during the school year had no detectible effects. Second, in Uttarakhand, where the volunteers were embedded by the school system, the program including teacher training, volunteers, and material had no effect whatsoever. This may be because teachers used volunteers as a substitute and exerted less effort. Third, the program in Haryana, which included an explicit component empowering government monitors to supervise the program, had strong effects on Hindi test scores: learning levels increased by an average of 0.15 standard deviations.

The remainder of this paper is organized as follows. Section 2 describes the programs and experimental design. Section 3 analyzes the program effects on test scores as well as the process monitoring data. Section 4 offers an interpretation of the results. Section 5 concludes.

2 Programs and Experimental Design

The results discussed here come from randomized evaluations in which a common pedagogy was implemented in different forms: different locations, with different levels of intensity, by different instructors (government-school teachers, Pratham staff, or community volunteers), sometimes at different times (during the school year, or school hours, other times after hours, or during the summer), sometimes in different physical locations (within the government school building, elsewhere in the community). In some cases the pedagogy was delivered successfully, in other cases unsuccessfully. Ultimately, the model evolved into the "Teaching at the Right Level" model, which has been most recently implemented in Punjab, Bihar, and Haryana, but was only rigorously evaluated in Haryana.

In this section, we first introduce the common elements of the Pratham programs before discussing how the specific context and implementation differed in each location.

2.1 Common Elements

2.1.1 Combined Activities for Maximized Learning (CAMaL) pedagogy

The core element of all Pratham's Programs discussed here is the pedagogy: CAMaL. It is meant for children who are behind their expected level of achievement, according to the "age-appropriate" curriculum used in government schools. It focuses on remedial skills–basic literacy and numeracy, building the competencies of "Listening, Speaking, Reading, Writing and Doing," It uses materials that have been designed by Pratham, and involves recitation and is highly interactive. The pedagogy has evolved based on Pratham's own experience, internal assessments, as well as external randomized evaluations, including the evaluations reported on here (see also Banerjee et al., 2007). Perhaps out of line with the amount of work and expertise that has gone into designing and refining the curriculum, it is relatively simple to administer. Instructors who use the pedagogy are often given only one week's training initially, with some monitored hands-on experience, and then a week or more of follow-up training over the course of the year.

2.1.2 Program Context

Within India's federal system, primary education is the responsibility of the states, the second tier of government. There was substantial Central Government finance of education via the Sarva Shiksha Abhiyan ("Education for All") program in the 2000s, that was passed through the State governments, albeit with parallel administrative arrangements, and in 2009, under the Right to Education Act.

The programs evaluated were all implemented in northern states, in which Hindi is the primary written and spoken language (and is also the primary language of roughly 40% of the population of the country). .

2.1.3 Program

Although the pedagogy has been introduced through various programs for over a decade, since 2007 it has been rolled out across India primarily as part of its flagship Read India Program. In all locations, the program used the CAMaL pedagogy, targeting primary school-aged children, focusing on remedial skills in communities and rural government primary schools. Pratham directly trained all instructors, whether the instructors were to be Pratham staff, community-based volunteers, or government-employed. In all cases, the programs were intended to be a complement to, and not a substitute for, the formal educational curriculum that is established in each State.

Finance has came from a variety of sources, including government budgets and various grant financing, including the grants from the Bill & Melinda Gates Foundation and The William and Flora Hewlett Foundation.

2.1.4 Evaluations

In each of the examples discussed in this paper, the program was evaluated using a randomized evaluation. Within the randomization design, the interventions were compared to a control group. In some cases the pedagogy was implemented in different forms within the same context and time, and the different implementation strategies were compared and evaluated within the same study design.

For the new evaluations reported here, the effectiveness of all interventions are measured by childrens' performance on the ASER language and math tests. The ASER reading test measures child's reading ability in terms of the following classifications: beginner (cannot recognize letters), letter, word, paragraph, and story level. The ASER Math test measures child's math level in terms of the following classifications: beginner (can not identify single-digit numbers), numbers 1-9, numbers 11-99, subtraction, and division. These tests have been extensively piloted by Pratham as well as researchers at the ASER Centre and J-PAL .

2.2 First Pilot of Read India in Uttar Pradesh

Pratham's early pilot of Read India–the "Learn to Read" program–in Uttar Pradesh (UP) came on the heels of a randomized evaluation of Pratham's *Balsakhi* program. The Balsakhi program was a remedial education program that used a pedagogy very similar to CAMaL, was implemented during school hours, in the school building, by paid para-teachers employed by Pratham, and in urban ares (Vadodara and Mumbai) (Banerjee et al., 2007). We learned from the Balsakhi program that a pedagogy which refocused teaching on remedial skills rather than the official school curriculum could have large and statistically significant positive impacts on student learning. The scalability of this model, however, was limited by (a) the fact that it was designed to be implemented in urban centers, whereas 70% of the Indian population is in rural areas, and (b) the fact that it relied on paying para-teachers, which was deemed unsustainable given the vast number of para-teachers required to reach all children in need. The UP model used the CAMaL pedagogy, but, in contrast was implemented in rural areas instead of urban; it was administered through the communities directly rather than in partnership with the formal education system; and it was implemented by recruited volunteers rather than paid Pratham teachers. The Read India program was being piloted in many districts in UP—not just in the study areas—as well as in Bihar. Therefore the evaluation is of an already massively-scaled-up program, especially for a pilot.

2.2.1 Context

The UP program evaluation took place in 2005-6 in Jaunpur district. According to the 2001 census, UP had a population of 166 million, making it the most populous state in India. At the time of the study, it also had among the lowest levels of child learning in India: 49% of children in grade 5 could not read at a 2nd-grade level (ASER Centre, 2005). The district of Jaunpur was chosen for the study because it had average levels of literacy for the state, and because there was no prior Pratham presence in the district. The 280 study villages were randomly selected from a random subset of four blocks (sub-district administrative units) within Jaunpur district.

2.2.2 Program details

The UP program was community-driven, and part of a larger study of community participation and mobilization in education. On the shoulders of the community mobilization program, Pratham staff recruited volunteers from the communities where classes were to be held, and the staff then trained and monitored the volunteers during the intervention period. For the most part, it did not rely on government school personnel or resources (see Banerjee et al., 2010). Classes were held after school hours during the 2005-2006 school year. Occasionally these classes were held in the school building; but more often they were held in some other location within the community.

2.2.3 Evaluation Design

Out of the sample of 280 villages in Jaunpur, 85 villages were randomly assigned to serve as the control group, and 65 were assigned to the volunteer intervention group. The remaining 130 villages were assigned to other interventions that focused on disseminating information and promoting civic participation, and are discussed and reported in Banerjee et al. (2010). Classes were run in 55 out of the 65 volunteer treatment villages.

Baseline data were collected in March and April of 2005. Thirty households were randomly selected in each village, and all children aged 7-14 were given ASER reading and math tests. A randomly-selected 1/3 of these households were administered a household survey in addition to the ASER tests. The endline was conducted in March and April of 2006.

Further details on program and evaluation design can be found in Banerjee et al., 2010.

2.2.4 Lessons

We learned from UP that the pedagogy could delivered on a much larger geographic scale in rural areas using community volunteers. However, take-up of the program was low: only 17% of eligible students participated in "Learn to Read" classes. Scaling up on the intensive margin—achieving higher take-up—would likely require working with the government education system, involving teachers directly. Pratham had begun to do so in Punjab, but it had not yet been rigorously evaluated

. Simultaneously, Pratham refined the Read India program—incorporating several elements of community mobilization used the UP pilot—and massively scaled-up in 19 states, covering 33 million children in the 2008-2009 school year .

2.3 Bihar and Uttarakhand: Read India Scale-up

The evaluation in Bihar and Uttarakhand was the first assessment of the fully defined Read India Program, which was being implemented in partnership with the state government school system.

2.3.1 Context

The Read India programs were studied over the two school years of 2008-09 and 2009-10. Two states were selected, in order to explore the potential heterogeneity in effects in different contexts. Bihar and Uttarakhand were chosen for their contrasting characteristics with regard to initial learning levels, socio-political contexts and administrative capabilities. Uttarakhand was somewhat more developed, had a lower out-of-school population and was slightly better state-wide test scores than Bihar. Bihar, in contrast, has a history of educational disadvantages, with one of the largest out-of-school populations of children at the beginning of the 1980s. However, in 2005 the government began a major push on education, with a more than doubling of the number of teachers and a major push to get primary-age children into school.According to ASER 2008, 5.7 percent of children in the 6-14 age group were out of school in Bihar, in 2008, compared with 1.2 percent in Uttarakhand (ASER Centre, 2009). Only 51 percent of Bihari children in grade 2 could read at word level, compared with a still-dismal 62 percent in Uttarakhand. Only 46 and 49 percent of grade 4 children could read a grade 2 text respectively in the two states.

A further difference between the two states lies in their history of interactions with Pratham. In Bihar, Pratham had developed a strong working relationship with the state education service, a product of intensive involvement in supporting the government in earlier campaigns to get children into schools While Pratham was already working in Uttarakhand in 2007, in Bihar there had been a much more intensive engagement in joint programs. These differences in working relationships

became compounded in implementation by differences in staff turnover: in Bihar, the top managers in both the education ministry and Pratham remained the same throughout the period evaluated. By contrast, in Uttarakhand, there were changes in both government and Pratham staff over the course of the implementation period. Last, in Bihar there had been no special activity of a Read India type in the project area. However, in Uttarakhand the state government had implemented an intervention called NEEV, using Pratham materials for basic skills (then known as Learning to Read), with some Pratham inputs throughout the state in the year before the project started.

We selected geographic areas that were broadly typical for each State in terms of socio-economic conditions, but that were also feasible for large-scale survey work. For Bihar, this led to the selection of blocks in West Champaran, a poor district on the border of Nepal, with levels of educational disadvantage around the average for rural Bihar, and, importantly for the survey work, not subject to regular flooding. For Uttarakhand, blocks in the lowland districts of Dehradun and Haridwar were selected; it would not have been feasible to manage a large survey in the mountainous areas in the time required. Once the spatial domain was defined, all villages were listed within the area, and then randomly assigned to the various treatments within the school year. For the summer camp in Bihar, all villages assigned to school-year treatments were assigned to receive summer camp classes. All children and all government schools within each village constituted the treatment groups. In Bihar, the study involved 158 villages with 264 government schools allocated equally to the three school-year treatment groups and control, while in Uttarakhand, there were 114 villages with 122 government schools, again allocated equally to the two school-year treatment groups plus control.

2.3.2 Program

The program was implemented in four different forms in direct coordination with schools in the treatment villages. In the first intervention, remedial training was provided to children in a one-month summer camp, run in school buildings by government school teachers, that occurred in June 2008 in Bihar. Pratham provided materials and training for government school teachers, and also

trained a volunteer who supported teachers in the classroom. The government school teachers were paid extra by the government for their service over the summer period.The other three interventions were conducted during the school year. The second intervention, also only implemented in Bihar, involved the distribution of Pratham materials with no additional training or support (referred to here as the M treatment). The third intervention involved provision of materials plus training of two teachers from every school in the Pratham pedagogy, followed by regular visits of Pratham field staff to support teachers and monitor their use of materials, organization of students and other aspects of the pedagogy (the TM treatment). The fourth intervention also involved materials and training for teachers, but this was supplemented by unpaid volunteers, who were selected from the village and trained separately by Pratham staff (the TMV treatment). Following local discussions with government education staff, there were distinct approaches to the use of the volunteers: in Bihar, volunteers worked outside school hours in the community, though they sometimes used school premises. In Uttarakhand, the model was for volunteers to support teachers in their work within the school day.

2.3.3 Evaluation Design

For the research design in Bihar, about 120 villages were randomly assigned to receive summer camps and 40 no summer camp–the control group. The effectiveness of summer camps was evaluated based on the comparison of these two groups. Subsequently, those that received summer camps were randomly assigned to receive one of the three remaining interventions: M, TM, or TMV with about 40 villages in each group. The summer-camp control group remained the control group during the school year. In Uttarakhand, none of the villages received summer camps, and were randomly assigned to the control group or TM or TMV, with about 40 villages in each group.

In both states, the control groups received all normal government support activities, but no additional support from Pratham during the project period.

To identify the sample, a census of households was conducted in each village in the study, and households with children aged 5 to 14 were randomly selected from this list to be tested and

surveyed. Total sample size was derived from power calculations using the impacts found in the Jaunpur study (Banerjee et al. 2007) as a benchmark. Overall some 5,200 households with 12,300 children were included in Bihar (about 33 households per village) and 4,050 households with 8,900 children in Uttarakhand (about 35 households per village). Households and children were surveyed and tested prior to the initiation of any intervention activity and at the end of each of the two school years. In Bihar, a sample of children from villages participating in the summer camp (and controls) were surveyed immediately after the camp.

In addition to the ASER tests described above, children were also administered written Hindi and math tests. These tests, each taking 20-25 minutes to administer, tested higher level competencies than the standard ASER tests. In Hindi, separate tests were used for children in grades 1-2 and 3-5.

All government schools in the study were also surveyed, with interviews of headmasters, teachers and observation of school conditions. School surveys were designed to be implemented at baseline, and during the first and second school-year endlines. In both school years, JPAL undertook unannounced visits to all schools to observe school conditions and track program implementation. These surveys involved collection of data on school attendance, self-reported material usage by teachers, interviews with children, teachers and volunteers, and observations of Hindi and math classes.

For the baseline, the Indian branch of an international survey company was selected for the survey work. Their performance was substantially below expectations. In Bihar, the survey was judged to be of just adequate quality to use the data, but baseline test scores were missing or had to be dropped for a significant number of children. For all subsequent surveys, the study switched to a local survey company, based in Patna, Bihar, with satisfactory results. For Uttarakhand, the survey team was worse, and evidence of forgery of data by surveyors was discovered. The initial baseline had to be completely thrown out, and subsequent surveys were self-managed. In order to get baseline information before the intervention started, a shorter survey was undertaken that only involved the ASER test of children. The "baseline" household survey was conducted as part of

the first year endline, and the full array of child tests were included in both first and second school year endlines. The "baseline" school surveys were also undertaken in the course of the 2008-09 school-year.

JPAL field staff observed overall implementation and training of teachers, and held regular discussions with Pratham field workers to obtain information on discussions with state and district governments, and form a qualitative impression of progress. This was complemented by additional qualitative work, including an "institutional review" undertaken by the Centre for Policy Research, New Delhi, to inform our understanding of the workings of Pratham and the nature of interactions with state teaching services. (see Kapur and Icaza, 2010). A second qualitative analysis was undertaken by JPAL field staff involving semi-structured interviews with teachers, focus groups of teachers in resource centres, and focus groups of mothers in a small sample of schools and villages in both Bihar and Uttarakhand (Sharma and Deshpande, 2010).

2.3.4 Lessons

The lessons from Bihar and Uttarakhand were two-fold. First, with Pratham support and training, the pedagogy could be delivered by government teachers when outside of the normal education system (i.e. summer camps). Second, when integrated into the school year, training teachers in the pedagogy and providing materials did not necessarily result in successful delivery of the pedagogy, and ultimately, there was no improved learning. The only models that worked were those delivered outside of the formal education system–either during the summer period, or by volunteers independent of the school. From the experience, a theory emerged that teachers were held accountable only for completion of the government-mandated curriculum, and that the additional directives from the state government and Pratham staff did very little to incentivize teachers to redirect focus elsewhere. Recognizing that working with government schools was the best, and perhaps only way to achieve scale, Pratham continued to explore models to integrate their pedagogy into the formal educational system, while confronting the challenge of teacher incentives head on.

2.4 Haryana: An evaluation of the TaRL model

Following an encouraging experience delivering Read India with the Punjab school system, Pratham refined a model of delivering their pedagogy that put a great emphasis on monitoring of teacher activities, and physical restructuring of the classrooms where students were reorganized by ability rather than by grade. This most recent incarnation of the model, now referred to as the "Teaching at the Right Level" (TaRL) model, is what was tested in Haryana.

2.4.1 Context

Haryana is more developed than other states in India. Haryana ranks third highest among Indian states in per capita income (Reserve Bank of India 2011). Similarly, at the time of our study's baseline, student learning levels were higher than the national average: 52 percent of students in classes 1 to 8 in Haryana could read a grade 2 level text while the corresponding number was 38 percent for India as a whole, placing Haryana, again, third amongst Indian states (ASER Centre 2011a). Levels of nutrition, on the other hand, do not make Haryana an outlier: the percentage of children under 3 classified as stunted, wasted and underweight is similar for Haryana and India as a whole (IIPS & Macro International 2008).

Two districts in Haryana were chosen for the evaluation: Mahendragarh and Kurukshetra. Lying at opposite ends of the state, these two districts were selected by GoH and J-PAL because of their different educational and economic profiles. Along basic measures of child literacy and numeracy, Kurukshetra falls primarily below the state average, while Mahendragarh is above the state average. 0.5% of children ages 6-14 in Mahendragarh were not enrolled in school in 2010, compared with 0.1% of children in Mahendragarh (ASER Centre 2011b). Kurukshetra is also more economically developed than Mahendragarh. For example, 96% of households in Kurukshetra are electrified, compared with 80% in Mahendragarh (Census of India 2011). Mahendragarh is also one of the two Haryana districts covered by the Backward Regions Grant Fund, a national program supporting India's least developed districts (NIRD 2009).

2.4.2 Programs

As implemented in Haryana, the pedagogy was implemented by government school teachers, within the school building, during the school day. Teacher training was conducted by Pratham staff. As with other Read India programs, teachers were trained to administer a brief (around five minutes per child) oral assessment of each student's reading ability in Hindi at the beginning of the school year. To send the a signal from above that teachers should teach at the level of the child (i.e. focus on remedial skills), contradicting the competing signal that they should complete the curriculum assigned for that grade, three innovations were included in the model.

First, training of teachers in the pedagogy was organized by the state government, and from the announcement, to the location and timing, all efforts were made to suggest that this was a Government of Haryana training, and not a Pratham training.

Second, a system of monitoring was set up that utilized government staff to monitor compliance with the TaRL model. The status quo monitoring system included an established process of school monitoring with block and district supervisors and field-level monitors, known as Associate Block Resource Coordinators (ABRCs). However, while the administration had provided general guidelines on the roles and responsibilities of the monitors and supervisors, no specific training on how to operationalize these responsibilities had been provided. ABRCs had also been used generally as "couriers": they collected information regarding various programs in the schools, delivered letters, etc., and were therefore effectively not utilized as program supervisors. Overall, the emphasis with respect to program/project management had been restricted to ad-hoc data collection on physical inputs, with no attention paid to outputs, let alone educational outcomes or impact.

ABRCs were trained on to implement the TaRL model , on how to monitor the intervention with observations, forms and basic statistical analysis, and on how to mentor teachers to ensure the model's fidelity. This was intended to send a signal from the government school system that teachers were institutionally accountable for implementing the TaRL model.

The third innovation, was that all students in each program school were, for part of the school day, to be reassigned and physically moved from their grade-based classrooms to classrooms based

on levels determined by the baseline assessment: Beginner (could not identify letters), Letter (could identify individual letters), Word (could read isolated words), Paragraph (could read brief sentences), and Story (could read longer paragraphs telling a story). Once classes were restructured into these level-based groups or "mahals," teachers were to use the mahal-specific part the CAMaL pedagogy. Classroom restructuring of this type had been part of the teacher training in Bihar and Uttarakhand, but it received more emphasis in the Haryana training. In addition, Beginning in the 2011-12 school year, the Government of Haryana (GoH) mandated that all schools add an extra hour of instruction to the school day. Within TaRL schools, the extra hour was to be used for class reorganization and teaching remedial classes using the Pratham curriculum. In the control schools, the extra time was used to conduct classes using their standard pedagogy. Restructuring the classroom simultaneously sent a signal that the intervention was government mandated, broke the status quo inertia of routinely following the curriculum, and made it practically easier for teachers to teach at the child's level.

2.4.3 Evaluation Design

Four blocks, two in each of Kurukshetra and Mahendragarh districts, were selected at random as our intervention sites. Across these four blocks, a total of 400 drawn from a list of all 467 government schools in the four blocks.

The study of the TaRL model was conducted alongside an evaluation of another government program: Continuous and Comprehensive Evaluation (CCE) . The 400 schools were assigned in equal proportions to either a control group, TaRL only, CCE only, and TaRL combined with CCE. The results of the CCE program are reported in Berry, et. al. (2014). We find no detectible effect of the CCE program. In the analysis that follows, we include dummies for assignment to the CCE program.

A key concern in our random assignment of schools to these treatment groups was that of spillovers between neighboring schools. In our study areas, schools of different levels may share the same grounds or even the same building. Due to fears of spillovers between schools sharing

a campus, randomization was conducted on the level of the school campus: a group of schools at different levels with the same name in the same locality, usually occupying a single building or complex of buildings.

Baseline testing took place in the 2011-12 school year, before implementation of the programs, and endline testing took place at the end of the 2012-13 school year, following implementation in schools assigned to the treatment groups. Local staff hired and trained by J-PAL South Asia administered and scored all tests.

In primary schools, our sample consisted of students who were in grades 1-4 at baseline. Tests were administered to up to 10 randomly selected students in each grade in each school (of an average of 17 students per grade) at both baseline and endline, yielding a sample of 12,576 students. For all students in the sample, we collected basic demographic data – including gender, caste, religion, age, and parents' occupations – as well as records of recent school attendance from school registers in each round of testing.

Primary school students in grade 3 or higher at either round of testing were also administered written Hindi and math assessments. The written tests, developed by J-PAL South Asia and Pratham for the evaluations in Bihar and Uttarakhand, test students on competencies which they should be able to demonstrate by the end of grade 4. The written Hindi and math tests were each scored on a scale from 1 to 12.5 at both baseline and endline.

In addition to student testing, data on school composition and teaching and evaluation practices were collected through surveys of school headmasters conducted simultaneously with student testing at both baseline and endline. The endline headmaster survey also included modules on program implementation, knowledge of how to implement the Pratham pedagogy, and involvement of ABRCs.

Finally, we incorporated an extensive program of process monitoring into our study design. This monitoring consisted of surprise visits to each of the 500 schools included in the evaluation by trained J-PAL monitors between August 2012 and March 2013. During these visits, monitors administered an extensive questionnaire that included modules on program implementation, the

availability of learning inputs such as textbooks and uniforms, monitoring by ABRCs, and other topics. Monitors also observed a randomly selected teacher for thirty minutes to collect data on teaching and evaluation practices in the classroom.

2.5 Descriptive Statistics

In each location, our analysis sample consists of the group of children targeted by the program. Because the programs targeted slightly different groups of children, our samples differ across locations. The Uttar Pradesh program targeted children in the communities aged 7-14. The Bihar summer camps targeted children who had completed grades 1-5 the year prior to the camps. The school-year programs in Bihar and Uttarakhand focused on children in grades 1-5. Because these programs ran for two years, we restrict the analysis sample to children who were in grades 1-5 in both years. The Haryana program targeted children in grades 2-5 in school. In Appendix A, we show that the main results are virtually unchanged when a consistent selection criteria is used across locations.

Table 1 displays basic summary statistics of key baseline variables for the control group across the four intervention areas. Several differences are worth noting. First, the Uttar Pradesh sample is older, at a higher grade level, and has higher test scores, owing to the different age selection criteria for surveyed households. Average age, grade level, and test scores are similar across Bihar, Uttarakhand, and Haryana. Second, because the Uttar Pradesh intervention was the only one to target children in the communities, only that sample includes a substantial number of out-of-school children.

Table 2 examines the balance of these same baseline variables across treatment and control groups. The Uttar Pradesh, Uttarakhand, and Haryana samples show few significant differences across these variables. On the other hand, in the Bihar sample, for 4 out of 7 variables, the F-test rejects equivalence between treatment and control groups at the 5% level.

3 Results

Table 3 presents the main results of the interventions on test scores. Overall, we find significant impacts on language test scores for the volunteer-only intervention in Uttar Pradesh, the TMV intervention in Bihar, and the TM intervention in Haryana. The TMV interventions in Bihar also had significant impacts on math scores, but neither of the other cases did so (and note that math was not included in the Haryana treatment.)

Our results are based on the following specification:

$$y_{ist} = \alpha + T_s\beta + \delta y_{ist_0} + X_{is}\theta + \varepsilon_{ist} \tag{1}$$

where y_{ist} represents the normalized test score for student i in village/school s at endline t (normalized by subtracting the baseline mean in the control group, and dividing by the baseline standard deviation),T_s is a vector of dummy variables indicating inclusion of the village/school in each treatment group, y_{ist_0}represents the student's normalized baseline test score, and X_{is} is a vector of student- and school-level characteristics common to all datasets, including gender, age, and grade. Standard errors are clustered at the level of village/school.[1] We run Equation (1) separately for each state, and separately for the Bihar summer camp and school year interventions.

The Bihar and Uttarakhand samples also include children who did not take the baseline test but were added based on the household surveys at midline or endline. We include dummy variables to indicate missing values for their baseline test scores.

3.1 Main Results

Columns 1 and 4 of Table 3 display the treatment effects of each intervention in each state on normalized ASER language and math scores. Panel A displays the results from Uttar Pradesh. The volunteer-only intervention resulted in a 0.05 standard deviation increase in language scores. However, only 8% of the children surveyed attended the classes. Using assignment to the treatment

[1] In Haryana standard errors are clustered by school campus, the level of randomization.

group as an instrument for takeup (and assuming no spillovers to children who did not attend), the IV estimate of the effect of the classes on children who attended equals 0.55 standard deviations. The program's effect on math scores is small and statistically insignificant.

Panel B displays the results of the summer camps in Bihar. The program had significant impacts on language and math of 0.09 and 0.10standard deviations, respectively. As in Uttar Pradesh, only a fraction—23%—of children attended the summer camp classes. The IV estimates of the impact of attending the summer camp on language and math is therefore 0.40 and 0.44 standard deviations, respectively.

Turning to the Bihar school year interventions, the only intervention that resulted in a statistically significant increase in test scores was the combined TMV intervention, resulting in a 0.13 standard deviation increase in language scores, and a 0.10 standard deviation increase in math. The TM and M interventions did not have statistically significant impacts on learning. For these interventions, confidence intervals allow us to reject effect sizes as large as the point estimate for the TMV intervention. For example, the 95% confidence interval for the effect on language allows us to reject effect sizes of over 0.12 standard deviations for training-materials, and over 0.08 standard deviations for materials only.

In Uttarakhand, neither the TMV nor the TM interventions resulted in statistically significant increases in test language or math scores. For language, the 95% confidence intervals allow us to rule out effect sizes of over 0.07 standard deviations for TMV, and 0.14 standard deviations for TM.

The Haryana TM intervention resulted in a statistically significant 0.15 standard deviation increase in language test scores. For math, the effect is small and not statistically significant. Here, the estimate is more precise than those for other interventions: the confidence interval allows us to reject effects larger than 0.03 standard deviations in math. This lack of results in math in Haryana is not surprising, given that the TaRL curriculum in that state included only Hindi.

3.2 Heterogeneity by Pretest Score

To examine whether the interventions were most effective for the highest- or lowest-performing children, we disaggregate impacts by baseline ASER level. Table 4 displays treatment effects (in standard deviations) for each baseline level. The first three columns display impacts on language scores. Across the four interventions with statistically significant impacts in the aggregate—Uttar Pradesh, Bihar Summer Camp, Bihar TMV, and Haryana TM—there are no consistent patterns of increasing or decreasing impacts by baseline test score. Instead, the largest impacts occurred for children at the letter level at baseline.

Columns 4 through 6 of Table 4 display the impact estimates for math, by baseline math level. Again, the two interventions with significant impacts—Bihar Summer Camp and Bihar TMV—do not show increasing or decreasing impacts by baseline score. Both interventions had the largest impacts on children at the subtraction level (the second-highest level possible).

3.3 Process Data

To shed light on intensity of implementation of each program, we turn to process data collected by J-PAL enumerators. Process data were collected for the school-year interventions in Bihar, Uttarakhand, and Haryana. In all three states, each school was visited once during the year. Interviewers collected information on teacher and student attendance, and observed classroom teaching. In Bihar and Uttarakhand, enumerators observed 4th grade Hindi and math classes, spending about 2 hours on classroom observations at each visit. In Haryana, enumerators observed a randomly-selected class for 30 minutes. We focus on teacher attendance and classroom observations, with particular attention to the differences between Bihar/Uttarakhand and Haryana. The monitoring surveys in each state contained different questions about classroom practices, but we compare similar sets of questions across locations.

We first examine the extent to which teachers had been trained in the Pratham methodology. Table 5 displays the percentage of teachers trained across locations and treatment groups. In Bihar, 86-87% of the teachers in the TM and TMV groups had been trained, while only 1-5% were

trained in the Control and M groups. In Uttarakhand, however, teacher attendance in training was lower: only 28% of teachers in the TM group, and 45% of teachers in the TMV group received training, compared with 15% in the control group. Compliance with training was the highest in Haryana, where 95% of teachers in the treatment group, and virtually no teachers in the control group attended training.

We next turn to classroom observations. In Each state, enumerators observed whether Pratham materials were being used during the observed classes. Table 6 displays the percentage of observed classrooms which were observed using Pratham materials. In both Bihar and Uttarakhand treatment groups, substantial numbers of teachers used Pratham materials. Use was highest in the Bihar TM and TMV treatments, where 49-52% of teachers were using the materials. In Haryana, teachers used Pratham materials in 74% of the classes reserved for TaRL, and in virtually none of the classes in either the control group or in treatment classes outside of the TaRL time.

Finally, we examine the arrangement of students in each school. In general, Pratham methodology includes grouping of students by learning levels, rather than by grade in school. Table 7 displays the observed classroom arrangements in the three states. Here the differences between Bihar/Uttarakhand and Haryana are the most stark: in very few cases (10 percent or less) in Bihar and Uttarakhand were students grouped by learning levels, even in the TM and TMV treatments. In Haryana on the other hand, over 90% of schools were grouped by learning levels during the time reserved for TaRL. At other times in TM schools, and in the control group, virtually no students were grouped by learning level.

In sum, we observe substantial numbers of trained teachers in Bihar, Uttarakhand, and Haryana, and substantial use of Pratham materials in all three states. By contrast, the classroom arrangement was substantially different in Haryana compared with Bihar and Uttarakhand: students were much more likely to be grouped by levels. In the following section we utilize these results in providing an overall interpretation of the findings across project locations.

4 Interpretation

We offer here an interpretation for the patterns of results across the three field experiments. This has three elements. First, there is clear support for potential effectiveness for the pedagogy that Pratham developed in improving basic learning levels in both reading and math—if the teaching and materials are aligned with a child's initial learning level. This is the essence of TaRL. Second, this can be effectively implemented even by village-level volunteers without formal teaching training, and by existing government teachers, when they have received relatively short training in how to do this. However, third, achieving the alignment between pedagogy and initial learning levels requires an explicit organizational effort to ensure that children are assessed, grouped and actually taught at the right level; this will not occur automatically within the existing government school system, but can be achieved by making this a task of the front-line of supervision.

As noted in the introduction, the context for this interpretation is the very low quality of existing rural government schools (with only slightly better measures results in rural private schools, or indeed in the majority of urban schools). Equally relevant is the very low progress in learning levels as children advance through standards. For example, the median gains in test results in the Read India schools in Bihar and Uttarakhand was zero over the two years, consistent with both the repeated cross-section from the ASER surveys and other panel data (ASER Centre, 2013). Past studies of India's government education system has emphasized low effort of teachers, as evident in teacher absence from schools and low levels of observed teaching activities when in school (Kremer et al., 2005) The PROBE report concluded that this was not just a matter of shirking, but also associated with teachers feeling disempowered, with multiple obligations from government outside teaching (PROBE Team, 1999). Other work has argued that rural teachers are embedded in reciprocal relations with politicians, making the system resistant to change (Beteille, 2009). These results led many observers to be pessimistic about effecting change in the governmental system, absent radical shifts in governmental processes and teacher's contracts (e.g. Pritchett and Murgai, 2007). While we only found significant positive impacts in a subset of the treatments, the results here are more optimistic of the potential for change, if there is sufficient commitment

within state-level government education systems. We expand on the findings here.

The evidence on the potential effectiveness of Pratham's pedagogy, if actually applied to the student's learning level, comes from the sequence of positive treatment impacts. Pratham made some modifications to the pedagogy over the course of the experiments, but we do not see these as being critical to the results. In Jaunpur, Uttar Pradesh, the pedagogy focused on basic reading and math (then known as Learning to Read), and Pratham's instructors taught the volunteers to align the teaching material to the learning level of the students. In the Bihar Summer Camp, the target group was students at low learning levels (Letter and below), and the pedagogy again focused on this level. As seen above, there were significant average effects despite targeting errors, with some children of higher levels joining the camps. For the regular school year in the Read India program in Bihar and Uttarakhand, Pratham had a more ambitious pedagogic approach, that could take students beyond basic learning to grade level , but this had the more basic pedagogy of Learning to Read embedded within it. The Bihar TMV treatment involved identification of weaker children, who were then taught by volunteers trained in the basic pedagogy outside the school day, with the aim of bringing all children to basic reading and math levels. This was thus effectively the same design as in Uttar Pradesh, except with teachers identifying weaker students, as opposed to Pratham staff. Finally, in Haryana the pedagogy again focused on basic reading skills, and was only applied to Hindi, and not to math, with significant gains in learning.

The key challenge in the sequence of experiments was how to get the pedagogy aligned with the child's level. This proved much tougher than originally anticipated. In particular, the Read India organizational designs in Bihar and Uttarakhand were ineffective in improving learning within the school year. Teaching to the Right Level was part of the training of government teachers by Pratham staff, along with the advice to set aside time each day to use the pedagogy, but this was less sharply presented than in the later Haryana training, according to Pratham managers. It was unclear from the process monitoring if the problems in implementation were fundamentally an issue of motivation/effort of teachers or of classroom organization and management. Cluster Resource Centre Coordinators (CRCCs, the equivalent of ABRCs in Haryana) were part of the

training, but not given special functions. Qualitative evidence suggests CRCCs were often preoccupied with other things than ensuring effective teaching, let alone organizational arrangements in treatment schools; though younger ones in the second year at least showed more interest (Sharma and Deshpande, 2010).

The Bihar summer camp showed that the pedagogy could be effective within government classrooms when there was sufficient alignment with children's learning levels. However, this result was ambiguous on potential effectiveness in the school year, both because teachers who chose to join the camp received extra pay (so there could have been selection and additional incentive effects) and because Pratham also trained volunteers to work alongside the teachers in the classroom.

The Haryana result is thus crucial to the overall interpretation. The TM treatment was within the regular school year, with an explicit part of the training being the sorting of children by levels for the learning pedagogy, backed by a concerted effort by Pratham and J-PAL staff to give the ABRCs specific functions in implementation. This had two components. First, Pratham's staff ensured ABRCs did 15-20 day practice classes before the teachers got trained, with the program designed to be led by ABRCs. Second, J-PAL staff helped set up a monitoring system and taught the ABRCs how to monitor teaching activities, with the intention of monthly visits to ensure schools were implementing the TaRL treatment. J-PAL's own process monitoring suggests that this was quite effective. The frequency of ABRC visits to schools varied from month to month, but at the final process monitoring visit to each school, 80% of schools reported a visit from an ABRC in the previous 30 days. Of those who reported a visit, 77% said that the ABRC spent over an hour in the school, and 93% said that the ABRCs observed a class in progress in at least one visit.

Evidence for the different level of implementation in Haryana compared with Bihar and Uttarakhand is reflected in the grouping of students in classes, as shown in the previous section. Students in Haryana were much more likely to be grouped by learning level. This was not because of lack of awareness of use of the materials. As shown above, teacher training was largely successful in both Bihar and Uttarakhand, and substantial use of Pratham materials was observed. The only substantive difference observed with the Haryana case was the successful organization into

learning groups, for which the additional effort on training and monitoring of ABRCs appears to be the key intervention.

5 Conclusion

Tackling the dismal levels of basic learning in India's government schools has been an extraordinarily tough development problem. Learning levels have seemed stuck at low levels for the vast majority of children, despite substantially increased resources going into the sector, and significantly greater attention to basic education. The heightened profile of basic education was reflected in the Right to Education Act, that was eventually passed in August of 2009. But there is no evidence yet that this has led to improvements in learning levels. This could lead to the view that the government education system is a lost cause, with teachers putting in low effort, and state education services ineffective in converting the additional resources into better learning outcomes. While private schooling and tutoring has been growing rapidly, including in rural areas, this would be a troubling result. In most states, a majority of rural children, and especially those from poorer households, still attend government schools. India is at risk of condemning many, if not most, of its children to learning levels that leave them with totally inadequate skills for their future working lives.

The sequence of field experiments reported here both help explains why education quality has been such a resistant problem and shows the potential for effective improvements in learning even within the rural governmental schooling system. The NGO Pratham developed over several years in the 1990s pedagogies designed to bring children up to basic levels of learning and math, especially for children that have been left behind by the education system and are well below grade level. The experiments confirm that these are highly effective when properly implemented. The pedagogy can work with only modest outlays in training (of several days, not months) with village-level volunteers, typically with high school education and no formal teacher training, and, under the right conditions, with existing government teachers.

However, the key issues are getting the pedagogy aligned with the initial learning level of the child, and then making this the task at hand for the teacher or volunteer. This is what is really meant by Teaching at the Right Level. This was relatively easy to effect for volunteers outside school hours, provided there was ongoing supervision by Pratham staff, as shown in the results for the Uttar Pradesh experiment and the Bihar treatment involving village-level volunteers. It also was effected without great difficulty in the Bihar summer camp, with local government teachers working with volunteers within schools. What was much more difficult was changing practices within schools, with teachers failing to reorganize classes around learning levels in both Bihar and Uttarakhand. The key intervention, that was central to the Haryana design, was to take the front-line of education administration seriously, and make the Assistant Block Resource Coordinators (ABRCs) the key actors in the monitoring and support of teachers. This was the only treatment in all of the experiments that led to significant learning gains without the use of volunteers.

This result is potentially of great importance for future change in India's government schools. In the case of Haryana it occurred with extensive engagement of both Pratham and J-PAL staff in the training and in the design of the ABRCs' monitoring system. This raises the issue of whether the broader government education systems will choose to follow this approach, and be able to implement it. At the time of writing the verdict is out on this, and it is likely that different states will follow different trajectories. Ironically, in Haryana, the state education service displayed little interest in the results of the experiment, and indeed actually abolished the ABRC position in 2013. Nonetheless, and Pratham's links with other state governments may facilitate scale-up elsewhere, based on the Haryana experience.

References

Tahir Andrabi, Jishnu Das, Asim Ijaz Khwaja, Tara Vishwanath, and Tristan Zajonc. Learning and educational achievements in punjab schools (leaps): Insights to inform the education policy debate. *World Bank, Washington, DC*, 2007.

ASER Centre. *Annual Status of Education Report (Rural) 2005*. Pratham Resource Center: Mumbai, 2005.

ASER Centre. *Annual Status of Education Report (Rural) 2012*. Pratham Resource Center: Mumbai, 2013.

National Education Assessment. 2007 national education assessment. Technical report, Accra, Ghana: Ministry of Education, Science, and Sports, 2008.

Abhijit Banerjee, Shawn Cole, Esther Duflo, and Leigh Linden. Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3):1235–1264, 2007.

Abhijit V. Banerjee, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. Pitfalls of participatory programs: Evidence from a randomized evaluation in education in india. *American Economic Journal: Economic Policy*, 2(1):1–30, 2010.

Tara Beteille. *Absenteeism, Transfers and Patronage: The Political Economy of Teacher Labor Markets in India*. PhD thesis, Stanford University, 2009.

International Growth Centre. The impact of remedial education through teacher community assistants on the quality of education in ghana. URL `http://www.theigc.org/`. Accessed September 12, 2014.

Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101:1739–1744, 2011.

Shobhini Mukherji James Berry, Esther Duflo and Marc Shotland. A wide-angle view of learning: Evaluation of the cce and lep programmes in haryana. *Mimeo, J-PAL*, 2014.

Michael Kremer, Nazmul Chaudhury, F. Halsey Rogers, Karthik Muralidharan, and Jeffrey Hammer. Teacher absence in India: A snapshot. *Journal of the European Economic Association*, 3 (2-3):658–667, 2005.

Pratham. Read india. URL `http://www.pratham.org/M-18-3-Read-India.aspx`. Accessed August 8, 2014.

Lant Pritchett and Rinku Murgai. Teacher compensation: Can decentralization to local bodies take india from the perfect storm through troubled waters to clear sailing? In *India Policy Forum 2006-2007*, pages 123–178, New Delhi, 2007. Sage Publications India Pvt. Ltd.

Paribhasha Sharma and Anupama Deshpande. Teachers' perception of primary education and mothers' aspirations for their children—a qualitative study in bihar and uttarakhand. *Mimeo, J-PAL*, 2010.

The PROBE Team. *Public Report on Basic Education in India*. Oxford University Press, New Delhi, 1999.

Uwezo. Annual learning assessment report. URL `http://www.uwezo.net`. Accessed September 12, 2014.

Table 1: Summary Statistics

	Uttar Pradesh	Bihar	Uttarakand	Haryana
	(1)	(2)	(3)	(4)
Female (%)	47.14	53.24	43.75	50.61
	(49.92)	(49.91)	(49.63)	(50.00)
Age (years)	10.32	8.46	8.05	9.04
	(2.29)	(2.09)	(2.00)	(1.57)
Grade at baseline	4.39	4.22	2.77	2.57
	(2.44)	(13.82)	(2.31)	(1.11)
Out of school	0.07	0.02	0.00	0.00
	(0.25)	(0.15)	(0.00)	--
Baseline oral Hindi test score (out of 4)	2.41	1.12	1.37	1.63
	(1.53)	(1.38)	(1.45)	(1.41)
Baseline oral math test score (out of 4)	1.13	0.53	0.51	0.71
	(1.12)	(0.89)	(0.84)	(0.93)
Baseline written Hindi test score (out of 12.5)		4.49		4.12
		(2.93)		(2.62)
Baseline written math test score (out of 12.5)		4.42		5.13
		(3.41)		(3.41)

Note: Means are shown for the control group in each sample. Standard deviations in parentheses.

Table 2: Balance Check of Randomization

A. Uttar Pradesh

	Control	V	P-value (t-test)
Female (%)	47.14	48.25	0.324
	(49.92)	(49.98)	
Age (years)	10.32	10.41	0.082
	(2.29)	(2.28)	
Grade at baseline	4.39	4.51	0.178
	(2.44)	(2.42)	
Out of school	0.07	0.07	0.903
	(0.25)	(0.25)	
Baseline oral Hindi test score (out of 4)	2.41	2.46	0.309
	(1.53)	(1.54)	
Baseline oral math test score (out of 4)	1.13	1.16	0.438
	(1.12)	(1.14)	

B. Bihar

	Control	TMV	TM	M	P-value (F-test)
Female (%)	53.24	52.97	55.08	52.70	0.481
	(49.91)	(49.93)	(49.76)	(49.94)	
Age (years)	8.46	8.49	8.52	8.42	0.717
	(2.09)	(1.99)	(2.04)	(2.07)	
Grade at baseline	4.22	4.88	4.38	4.85	0.674
	(13.82)	(15.71)	(14.24)	(16.13)	
Out of school	0.02	0.02	0.02	0.03	0.883
	(0.15)	(0.14)	(0.14)	(0.16)	
Baseline oral Hindi test score (out of 4)	1.12	1.30	1.15	1.04	0.041
	(1.38)	(1.51)	(1.42)	(1.37)	
Baseline oral math test score (out of 4)	0.53	0.67	0.53	0.46	0.003
	(0.89)	(0.99)	(0.90)	(0.83)	
Baseline written Hindi test score (out of 12.5)	4.49	5.04	4.50	4.51	0.174
	(2.93)	(2.85)	(2.79)	(2.74)	
Baseline written math test score (out of 12.5)	4.42	5.32	4.75	4.65	0.025
	(3.41)	(3.20)	(3.32)	(3.10)	

C. Uttarakand

	Control	TMV	TM	P-value (F-test)
Female (%)	43.75	43.63	44.53	0.895
	(49.63)	(49.61)	(49.72)	
Age (years)	8.05	8.01	8.05	0.908
	(2.00)	(1.96)	(1.96)	
Grade at baseline	2.77	2.63	2.49	0.101
	(2.31)	(2.14)	(1.61)	
Out of school	0.00	0.00	0.00	.
	(0.00)	(0.00)	(0.00)	
Baseline oral Hindi test score (out of 4)	1.37	1.38	1.37	0.995
	(1.45)	(1.44)	(1.43)	
Baseline oral math test score (out of 4)	0.51	0.50	0.47	0.840
	(0.84)	(0.81)	(0.78)	

D. Haryana

	Control	TM	P-value (t-test)
Female (%)	50.61	50.82	0.992
	(50.00)	(50.00)	
Age (years)	9.04	9.08	0.399
	(1.57)	(1.60)	
Grade at baseline	2.57	2.55	0.966
	(1.11)	(1.11)	
Baseline oral Hindi test score (out of 4)	1.63	1.63	0.742
	(1.41)	(1.41)	
Baseline oral math test score (out of 4)	0.71	0.68	0.922
	(0.93)	(0.91)	
Baseline written Hindi test score (out of 12.5)	4.12	4.00	0.569
	(2.62)	(2.61)	
Baseline written math test score (out of 12.5)	5.13	4.88	0.158
	(3.41)	(3.32)	

Note: Standard deviations in parentheses.

Table 3: Main Results		
	Language	Math
	(1)	(2)
A. Uttar Pradesh		
V	0.0463***	0.0142
	(0.0177)	(0.0195)
Observations	15612	15595
Attended reading class (2SLS)	0.546***	0.0639
	(0.201)	(0.223)
Observations	15576	15563
B. Bihar (Summer Camps, Midline)		
TMV	0.0945**	0.0980**
	(0.0384)	(0.0472)
Observations	2839	2838
Attended camp (2SLS)	0.404**	0.439**
	(0.161)	(0.196)
Observations	2685	2684
C. Bihar (School Year, Endline)		
TMV	0.127***	0.0994***
	(0.0349)	(0.0361)
TM	0.0517	0.0174
	(0.0371)	(0.0374)
M	0.00888	0.0363
	(0.0375)	(0.0353)
Observations	6490	6490
D. Uttarakhand (School Year, Endline):		
TMV	0.00940	0.0265
	(0.0325)	(0.0455)
TM	0.0617	0.0677
	(0.0410)	(0.0459)
Observations	3763	3762
E. Haryana:		
TM	0.150***	-0.00535
	(0.0174)	(0.0168)
Observations	11963	11962

Notes: Standard errors in parentheses (clustered at level of randomization). Regressions control for baseline test scores, as well as gender, age, and standard at baseline. Test scores are normalized using the mean and standard deviation for the control group in each test's respective round. *Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

Table 4: ASER Results by Baseline ASER Score

	Language				Math			
	Nothing	Letter	Word/Para.	Story	Nothing	Number	Subtraction	Division
A. Uttar Pradesh								
V	0.0417	0.121***	0.0663**	0.0109	-0.0166	-0.00105	0.0387	0.0780**
	(0.0370)	(0.0414)	(0.0322)	(0.0132)	(0.0243)	(0.0380)	(0.0451)	(0.0312)
Observations	2289	3541	3673	6109	5967	4534	2125	2969
B. Bihar (Summer Camps, Midline)								
TMV	0.0803	0.157**	0.0120	0.00509	0.0650	-0.0483	0.217**	0.145*
	(0.0676)	(0.0639)	(0.0641)	(0.0431)	(0.0625)	(0.114)	(0.0960)	(0.0831)
Observations	418	459	503	582	816	383	443	318
C. Bihar (School Year, Endline)								
TMV	0.129**	0.172**	0.0966	0.0158	0.0715	0.0904	0.0943	0.0871
	(0.0550)	(0.0795)	(0.0643)	(0.0628)	(0.0462)	(0.0843)	(0.0945)	(0.0840)
TM	0.0650	0.141	0.00830	-0.0389	0.0383	0.0324	0.0526	0.0221
	(0.0466)	(0.0882)	(0.0706)	(0.0880)	(0.0418)	(0.0967)	(0.118)	(0.0860)
M	-0.00599	0.0601	-0.0151	0.0159	0.0547	-0.0267	0.104	-0.0428
	(0.0484)	(0.0780)	(0.0668)	(0.0649)	(0.0391)	(0.0872)	(0.104)	(0.120)
Observations	2174	1006	764	556	3067	622	546	248
D. Uttarakhand (School Year, Endline):								
TMV	-0.0694	0.0475	0.00329	-0.0242	-0.0224	0.152*	0.00413	-0.0164
	(0.0533)	(0.0550)	(0.0486)	(0.0197)	(0.0448)	(0.0907)	(0.0885)	(0.0825)
TM	0.0661	0.0661	0.0575	-0.0176	0.0707	0.135	0.0665	0.0185
	(0.0637)	(0.0637)	(0.0547)	(0.0141)	(0.0546)	(0.0951)	(0.0925)	(0.0969)
Observations	1285	1100	631	535	2373	747	288	143
E. Haryana:								
TM	0.131***	0.213***	0.144***	0.0388**	-0.00749	0.00556	-0.0628	-0.00503
	(0.0335)	(0.0289)	(0.0249)	(0.0166)	(0.0213)	(0.0314)	(0.0405)	(0.0372)
Observations	3162	3627	3168	1919	6720	3009	1371	770

Standard errors in parentheses (clustered at level of randomization). Regressions control for baseline test scores, as well as gender, age, and standard at baseline. Test scores are normalized using the mean and standard deviation for the control group in each test's respective round.
*Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

Table 5: Teacher Training Attendance (School-Level Means)

A. Bihar School Year	Control	M	TM	TMV
Teachers attended Pratham training (%)	1.48	5.6	86.32	86.94
Observations	63	60	64	65
B. Uttarakhand	Control	TM	TMV	
Teachers attended Pratham training (%)	15.24	27.5	44.87	
Observations	41	40	39	
C. Haryana	Control	TM		
Teachers attended Pratham training (%)	0.35	94.71		
Observations	100	97		

Note: For Bihar and Uttarakhand, Pratham training includes only Std. 2 and Std. 4 Hindi and Math teachers; for Haryana, it includes all teachers in each school.

Table 6: Use of Selected Pratham Materials, by Treatment Group
(percent of classrooms observed in which material was used)

A. Bihar School Year	Control	M	TM	TMV
Letter flashcard	0	1.7	22	14.5
Punctuation flashcard	0	17.6	22.7	23.4
Storytelling flashcard	0	5.9	6.1	12.9
Vocabulary flashcard	0	3.4	26.5	29
Other Pratham materials	0	4.2	4.5	8.9
Any Pratham materials	0	24.4	49.2	52.4
Observations	109	119	132	124

B. Uttarakhand	Control	TM	TMV
Letter flashcard	0	2.7	5.4
Punctuation flashcard	0	9.5	18.9
Storytelling flashcard	0.6	8.1	8.1
Vocabulary flashcard	0.6	6.1	10.8
Any Pratham materials	1.3	13.5	23
Observations	156	148	74

C. Haryana	Control	TM (TaRL time)	TM (Non-TaRL time)
Letter chart	1	25	0
Story card/Word card/Letter card	0	9.5	0
Story books	0.0	69.0	2.0
Any Pratham materials	1.0	73.8	1.9
Observations	198	84	106

Table 7: Arrangement of Children in Classrooms, by Treatment Group (Percent)

A. Bihar School Year	Control	M	TM	TMV
Hindi				
Smaller groups by standard	20.7	31.1	14.8	17.3
Smaller groups by learning levels	0	1	4.6	0
One big group	79.3	68	80.6	82.7
Observations	82	103	108	110

B. Uttarakhand	Control	TM	TMV
Hindi			
Smaller groups by standard	54.2	49.2	54.4
Smaller groups by learning levels	10.2	6.6	1.8
One big group	35.6	44.3	43.9
Observations	59	61	57

C. Haryana	Control	TM (TaRL time)	TM (Non-TaRL time)
Taught by mahals	0.0	91.7	2.0
Other arrangement	100.0	8.3	98.0
Observations	193	84	100

A Appendix: Main Results with Restricted Sample

The programs in each location focused on a slightly different target group, and as such our analysis samples differ slightly by age and grade. Appendix Table 2 re-creates the main results of Table 3 using a consistent selection criterion across locations. The most restrictive criterion is that of Haryana, which focused on grades 2-5 during the program year. We therefore restrict the Uttar Pradesh sample to children grades 2-5, and the Bihar and Uttarakhand samples to students in grades 2-5 in both years of the program.

As shown in Appendix Table 1, the Uttar Pradesh, Bihar, and Uttarakhand results are virtually unchanged when using the more restrictive selection criterion. The Uttar Pradesh, Bihar summer camp, and Bihar TMV results have similar magnitudes and remain statistically significant. This analysis provides evidence that the differences in results across locations is not a result of different ages and grades used in the main analysis.

B Impacts on Written Test Scores

Only the Bihar, Uttarakhand, and Haryana interventions were evaluated using written tests. In Bihar and Uttarakhand, written tests were administered to grades 1-5, while in Haryana, written tests were administered for grades 3-5 only. As a result, we split the analysis into grades 1-2 and grades 3-5.

Columns 2-3 of Appendix Table 2 display the results for language, and Columns 4-5 display the results for math. The results consistently mirror the results for the oral ASER tests: the TMV intervention in Bihar had statistically significant impacts on both math and language, while the TM intervention in Haryana had a statistically significant impact on language. We also note that the estimate for the materials-only intervention in Bihar was associated with an increase in test scores by 0.08 standard deviations, a result significant at the 10 percent level. However, there were no other significant impacts of this intervention on either language or math scores, and as such we note that this result may have been due to chance.

Appendix Table 1: Main Results (Impact of Treatment at Endline)
Restricted Sample

	Language	Math
	ASER	ASER
A. Uttar Pradesh		
V	0.0514*	0.0238
	(0.0271)	(0.0246)
Observations	7848	7839
Attended reading class (2SLS)	0.435**	0.122
	(0.185)	(0.170)
Observations	7836	7827
B. Bihar (Summer Camps, Midline)		
TMV	0.0931**	0.0859*
	(0.0412)	(0.0466)
Observations	2602	2602
Attended camp (2SLS)	0.361**	0.336*
	(0.162)	(0.181)
Observations	2413	2414
C. Bihar (School Year, Endline)		
TMV	0.103***	0.0814*
	(0.0383)	(0.0454)
TM	0.0674	0.0201
	(0.0494)	(0.0535)
M	0.0672*	0.0808*
	(0.0402)	(0.0470)
Observations	3366	3366
D. Uttarakhand (School Year, Endline):		
TMV	0.0227	0.0365
	(0.0343)	(0.0539)
TM	0.0660	0.107*
	(0.0454)	(0.0550)
Observations	2632	2631
E. Haryana:		
TM	0.150***	-0.00535
	(0.0174)	(0.0168)
Observations	11963	11962

Notes: Standard errors in parentheses (clustered at level of randomization). Regressions control for baseline test scores, as well as gender, age, and standard at baseline. Test scores are normalized using the mean and standard deviation for the control group in each test's respective round. *Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.

Appendix Table 2: Written Test Results (Impact of Treatment at Endline)						
		Language			Math	
	ASER	Written (Std. 1-2)	Written (Std. 3-5)	ASER	Written (Std. 1-2)	Written (Std. 3-5)
	(1)	(2)	(3)	(4)	(5)	(6)
C. Read India - Bihar (School Year, Endline)						
TMV	0.127***	0.0902*	0.149***	0.0994***	0.101*	0.121**
	(0.0349)	(0.0488)	(0.0512)	(0.0361)	(0.0515)	(0.0505)
TM	0.0517	0.0433	0.0579	0.0174	0.0256	0.0161
	(0.0371)	(0.0528)	(0.0631)	(0.0374)	(0.0571)	(0.0581)
M	0.00888	0.00342	0.0840*	0.0363	0.0353	0.0742
	(0.0375)	(0.0496)	(0.0493)	(0.0353)	(0.0571)	(0.0473)
Observations	6490	3406	3086	6490	3406	3085
D. Read India - Uttarakhand (School Year, Endline):						
TMV	0.00940	-0.0466	0.0131	0.0265	-0.0582	0.00970
	(0.0325)	(0.0703)	(0.0367)	(0.0455)	(0.0657)	(0.0414)
TM	0.0617	0.0988	0.0567	0.0677	0.0263	0.0234
	(0.0410)	(0.0759)	(0.0452)	(0.0459)	(0.0710)	(0.0395)
Observations	3763	1263	2501	3762	1263	2500
E. Haryana Education Project:						
RE-LEP	0.150***		0.134***	-0.00535		0.0193
	(0.0174)		(0.0209)	(0.0168)		(0.0224)
Observations	11963		9204	11962		9204

Standard errors in parentheses (clustered at level of randomization). Regressions control for baseline test scores, as well as gender, age, and standard at baseline. Test scores are normalized using the mean and standard deviation for the control group in each test's respective round.

*Significant at the 10 percent level. **Significant at the 5 percent level. ***Significant at the 1 percent level.