Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India

Abhijit Banerjee, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton*

August 2016

Abstract

Previous randomized studies have shown that addressing children's current learning gaps, rather than following an over-ambitious uniform curriculum, can lead to significant learning gains. In this study, we evaluate a series of efforts to scale up the NGO Pratham's approach to teaching children according to their actual learning level, in four Indian States. While this approach was previously shown to be extremely effective when implemented with community volunteers outside of school, the objective of these new scale-up evaluations was to develop a model that could be implemented within the government school system. In the first two instances (Bihar and Uttarakhand), the methodology was not adopted by government schoolteachers, despite well-received training sessions and Pratham support. Motivated by the quantitative and qualitative analysis of these early attempts, we adapted the approach and designed large-scale experiments in the states of Haryana and Uttar Pradesh to test two new scale-up models. In Haryana, teachers received support from government resource persons trained by Pratham, and implemented the approach during a dedicated hour. In Uttar Pradesh, Pratham volunteers implemented high-intensity, short-burst "learning camps" for 40 days, in school and during school hours, with additional 10-day summer camps. Both models proved effective, with gains in language of 0.15 standard deviation in Haryana, and 0.70 standard deviations in Uttar Pradesh, on all students enrolled in these schools at baseline. These two models provide blueprints that can be replicated inside other government systems.

*Banerjee: MIT, NBER, and BREAD, banerjee@mit.edu; Banerji: ASER Centre and Pratham, rukmini.banerji@pratham.org; Berry: Cornell University, jimberry@cornell.edu, Duflo: MIT, NBER, and BREAD, eduflo@mit.edu, Kannan: J-PAL, IFMR (Institute for Financial Management and Research), harini.kannan@ifmr.ac.in, Mukerji: J-PAL, IFMR (Institute for Financial Management and Research), shobhini.mukerji@ifmr.ac.in, Shotland: J-PAL, shotland@mit.edu, Walton: Harvard University, michael_walton@hks.harvard.edu. Thanks to Richard McDowell, Harris Eppsteiner, Madeline Duhon and Laura Stilwell for research assistance; to Tanmayata Bansal, Sugat Bajracharya, Anupama Deshpande, Blaise Gonda, John Firth, Christian Larroulet, Adrien Lorenceau, Jonathan Mazumdar, Manaswini Rao, Shweta Rajwade, Paribhasha Sharma, Joseph Shields, Zakaria Siddiqui, Yashas Vaidya, Melanie Wasserman and Jeremy Welsh-Loveman for field management; to Diva Dhar and Radhika Jain for supervision; and to Shaher Bhanu Vagh for the educational test design and analysis. Special thanks to the staff of 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. We acknowledge that Rukmini Banerji is the CEO of Pratham, which represents a conflict of interest. She was involved in the design of the program and the evaluation, the implementation of the program, and the final paper review, but not with the data collection and the analysis of the data.

1 Introduction

Despite rapid gains in school enrollment over the past several decades, learning levels are stagnating in many developing countries. In India, for example, the 2005 ASER survey found that 39 percent of fifth graders could not read at a second-grade level (ASER Centre, 2005). Repeated surveys have found that Indian students perform significantly below grade-level standards in both math and reading, with essentially no improvements in any Indian states over the past several years (ASER Centre, 2014). Other large surveys in Pakistan, Kenya, Tanzania, Uganda and Ghana have found similar levels of children performing significantly below competency standards for core subjects, demonstrating the pervasiveness of this issue among developing countries (Andrabi et al., 2007; Uwezo, 2014; NEA, 2008).

These persistently low levels of achievement are not due to a lack of available solutions. In particular, a number of experimental studies have shown that simple changes in pedagogy can lead to significant improvements in learning levels. Interventions which focus on targeting teaching to the current learning levels of students, 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. However, while governments occasionally experiment with these interventions (IGC, 2014), most previous evaluation efforts have focused largely on interventions led by non-governmental organizations (NGOs), which rely on NGO staff or volunteers for teaching with new methods. There has been less experimentation on how to induce similar pedagogical changes among existing government schoolteachers. This is a critical gap, since reforming the government school systems would allow these practices to reach a much larger number of children and more effectively utilize the time they already spend in school.

This paper reports on our iterative efforts to design, experiment, re-design, and experiment again until two successful, replicable models were developed to scale up such a pedagogical

approach within a government school system in India. Over the last fifteen years, Pratham, one of India's largest NGOs, has been developing instructional models that can improve children's learning at scale in a cost effective and simple way. In Pratham's approach, instructors teach basic language and mathematics according to the level of the children, rather than a prescribed age-grade syllabus. Children are organized into groups by ability levels and are taught using level-appropriate teaching, learning activities, and materials. Learning goals are simple and clear, and ongoing measurement is used to assess progress. A previous study (Banerjee et al., 2010) showed that, when implemented with carefully monitored but lightly trained community volunteers in the state of Uttar Pradesh, the program (which focused on reading) significantly improved the reading ability of children who attended the program.

To test whether Pratham's model could be adapted and scaled within the government system, we implemented several different experiments in the states of Bihar, Uttarakhand, Haryana, and Uttar Pradesh between 2008 and 2014. In the first experiment, the "summer camp" (implemented in Bihar during the summer of 2008), government teachers, oriented by Pratham and assisted by village volunteers, were recruited to provide remedial education during summer vacation. These summer camps increased test scores by 0.07-0.09 standard deviations, compared to the control group, suggesting that teachers are capable of effectively teaching remedial skills to children.

Then, in two parallel experiments conducted in Bihar and Uttarakhand between 2008-2010, three arms were compared: (1) Some schools received learning materials without any additional support; (2) Teachers were trained to teach with the Pratham method during the school year, and received the appropriate materials; (3) Materials were provided, and both teachers and volunteers were trained (although volunteers had somewhat different roles in each state: In Bihar, they worked outside of schools, following the original Pratham model; whereas, in Uttarakhand, they were placed inside schools). Villages or schools were randomly selected to receive one of the three

models (or to be a control group), and data was collected at baseline and after two years. These experiments confirmed that an exact replication of the original Pratham volunteer model (which was implemented in Bihar) was effective in this different context, but also found that all the other scale-up models failed. This failure was due to several reasons: providing only materials was insufficient; trained teachers did not adopt the methodology and instead used the textbooks prescribed for the relevant grades, and when volunteers were placed inside schools, they were used by teachers as assistants to implement traditional methods.

Motivated by these results, process data from the research studies, considerable field experience, and qualitative interviews, Pratham developed two approaches to mainstream the program within government schools, which addressed the main gaps of the previous attempts: (1) The first, meant to be implemented in an environment with good teaching resources, relied on teachers to implement the program; however, it also made sure that teachers had a dedicated time in the day devoted to the program, and were supported from within the government hierarchy; (2) The second, meant to be implemented in very poor teaching environments, relied on Pratham staff, replicating the "camp" approach within the government schools.

The first of these two models was implemented by the state government in Haryana during 2012-2013. In this model, a cadre of government officials at the cluster level were first trained by Pratham with each cluster level official being responsible for providing academic support to 12-15 schools. Next, cluster officials conducted about 15 days of "practice" where they themselves taught daily in schools, based on the teaching-by-level approach that they had been trained in. Once the "practice" period was over, cluster officials rained teachers, with Pratham staff assisting as needed. Appropriate teaching materials developed by Pratham were used in the schools, and the program was implemented during a dedicated hour each day. In Haryana, the focus of the learning improvement effort was on reading; math was not covered. In addition, schools were supervised

and monitored by the cluster officials.

The second model was evaluated in Uttar Pradesh during 2013-2014. Here Pratham staff and village volunteers ran "learning camps" (periodic bursts of intensive activity, forty days in total) in schools. These took place during the school day and in school premises, and were supplemented with a 10-day intensive burst of "summer camps".

In Haryana, the school-year intervention, which was led by government teachers and supported by government supervisors, increased reading scores by 0.15 standard deviations (as mentioned earlier, the program focused only on reading). In Uttar Pradesh, where the methodology was taught by Pratham volunteers during school time, test scores increased by a staggering 0.61-0.70 standard deviations (from a much lower base than Haryana) in both math and language.

Guided by process monitoring data generated by the study, our interpretation of the results is that success of the program relies on systematic implementation of the two key ingredients of Pratham's methodology: (1) grouping of children by initial learning level, and (2) focusing on skills appropriate to that level (which was very basic in several of the groups). In Pratham programs that are run out of school and by Pratham volunteers or employees (such as the original study, or the the program tested in Bihar), these components are systematically implemented properly and the program improved test scores.

However, mainstreaming these fundamental changes in pedagogy into the regular school curriculum is difficult without careful top-down support and monitoring. In teacher-led programs that ran during the school year in Bihar and Uttarakhand, classrooms were never re-organized around initial learning levels. By contrast, the teacher-led program in Haryana, which was implemented in a dedicated time slot, and included supervision by government monitors, led to successful reorganization of classrooms and therefore higher reading scores. In Uttarakhand, even volunteers were not able to implement the grouping methodology since they were used by teachers as

assistants to carry out their regular activities. By contrast, in Uttar Pradesh, during the "camp" days in Uttar Pradesh schools, Pratham volunteers were allowed to reorganize and regroup children for a few hours during the school day to implement the teaching-by-level methodology, leading to the largest impacts. In both Haryana and Uttar Pradesh, the on-site monitoring and ongoing support that teachers and instructors received as the intervention progressed were likely a strong factor of success.

2 Programs and Experimental Design

The results presented in this paper come from multiple randomized evaluations in which a core pedagogical approach was implemented at different points in time, with differences in delivery method (government school teachers, or Pratham staff or volunteers), duration (intensive camp mode, or daily instruction over the course of the school year), and location (in school or in the communities).

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 Context

In government primary schools, the curriculum consists of a prescribed syllabus for each grade, with automatic promotion through grade 8. Thus, children falling behind the expected levels according to the syllabus in early grades continue to progress through the system unchecked without ever being forced, or given the opportunity, to learn basic skills. Teachers also have strong incentives to target their instruction to the highest-performing students in their classes. As a result, by grade 5, 39% of children are unable to read at a grade 2 level, a number that has not improved since 2005 (ASER Centre, 2005; ASER Centre, 2014).4

2.1.2 Teaching at the Right Level (TaRL) pedagogy

The core element of all Pratham's Programs discussed here is the pedagogy: it is called Combined Activities for Maximized Learning (CAMaL), but is also referred to as "Teaching at the Right Level" (TaRL). We call it TaRL below. This pedagogy has evolved over the years from Pratham's own intensive experience, internal assessments, as well as external randomized evaluations, including Banerjee et al. (2010) and Banerjee et al. (2007).

Although this strategy can be used for all students, the approach has been designed primarily for those who have reached grades 3 to 5 but are well behind their expected level of achievement, according to the "grade-appropriate" curriculum used in government schools. Its approach focuses on building basic reading and arithmetic skills. Students are grouped by their actual (rather than expected) learning level, and frequently tested. The method also utilizes interactive and attractive materials that have been designed by Pratham and are constantly updated.

Although a great deal of experimentation and effort have gone into its development, the actual instructional methodology is simple to use. Instructors who use the pedagogy are often given only a week of training, with practice sessions built in. During the intervention, there is typically on-site mentoring and ongoing monitoring as well as periodic reviews, discussions, sharing of experiences and refresher training as needed.

Although the TaRL pedagogy has evolved over the last decade, since 2007 it has been rolled out across India primarily as part of Pratham's flagship Read India program. The majority of the locations covered by Read India have been rural schools and communities. In almost all cases, Pratham has directly trained all instructors, whether the instructors were Pratham team members, or community-based volunteers. For training of government teachers, usually Pratham worked with a team of master trainers from the government side. The Read India program intends to be a complement to, and not a substitute for, the typical activities taking place in government schools.

Pratham's early pilot of Read India took place in 2005-2006 in Uttar Pradesh. Pratham staff mobilized youth volunteers from the village communities where Read India classes were to be held, and then trained and monitored the volunteers during the intervention period. Classes were held after school hours during the 2005-2006 school year. The program was evaluated by randomly assigning 85 villages to a control group, and 65 villages to a Read India intervention group, and two other village mobilization groups that did not involve the program. The evaluation showed significant improvements in reading skills among children in Read India villages, compared to all the other groups (Banerjee et al., 2010).

The instrumental variables estimates of treatment on the treated also demonstrated large gains for children actually attending the classes, suggesting that the TaRL pedagogy, run by village volunteers, could deliver learning content effectively. However, the "take-up" of the program in the village was low, limiting the impact on the average child: only 17% of students with the lowest learning levels participated in classes. Effectively working within the government education system would make it possible to reach a much larger number of students who are already in school.

2.1.3 Evaluations

The programs evaluated in this paper were all implemented in northern states of India, in which Hindi is the primary written and spoken language. In each of the examples discussed in this paper, the program was evaluated using a randomized controlled trial. 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.

The main outcome measures are children's 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 recognition, word recognition, paragraph reading (grade 1 level text), and story (grade 2 level text). The ASER Math test measures child's math level in terms of the following classifications: beginner (cannot identify single-digit numbers), one-digit number recognition, two-digit number recognition, subtraction, and division. In each of these assessments, children are marked at the highest level that they can comfortably attain. These tests have been extensively piloted by Pratham as well as researchers at the ASER Centre and the Abdul Latif Jameel Poverty Action Lab (J-PAL).

2.2 Bihar and Uttarakhand: Summer Camps, Materials, Teacher training, and Volunteers

Based on the Uttar Pradesh evaluation (Banerjee et al., 2010) and other implementation experiences, Pratham's Read India program evolved to include schoolteachers as well as volunteers. By 2008, a number of state governments had begun collaborating with Pratham in running learning improvement interventions in schools. The evaluation in Bihar and Uttarakhand was undertaken to evaluate the new model involving schoolteachers. The program and accompanying evaluations were the result of a partnership involving Pratham, the respective state governments, and J-PAL. The evaluations took place simultaneously in Bihar and Uttarakhand over the two school years of 2008-09 and 2009-10.

2.2.1 Context

For these first set of interventions, Bihar and Uttarakhand were chosen due to their contrasting characteristics with regard to initial learning levels, socio-political contexts and administrative capabilities. Uttarakhand is slightly more developed. In 2008, Uttarakhand had a lower out-of-school population, higher private school enrollment, and slightly better state-wide test scores than Bihar (ASER Centre, 2009; ASER Centre, 2014). According to the 2008 ASER survey, 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, and only 51 percent of Bihar's children in grade 2 could read at word recognition

level, compared with 62 percent in Uttarakhand (ASER Centre, 2009). Within each state, we selected geographic areas that were broadly typical for each state in terms of socio-economic conditions and education levels, but that were also feasible for large-scale survey work. In Bihar, this led to the selection of blocks in West Champaran, a poor district on the border of Nepal, while in Uttarakhand, blocks in the lowland districts of Dehradun and Haridwar were selected..

2.2.2 Programs

The program was implemented in four different forms in direct coordination with schools in the treatment villages. In **Summer Camps** (run only in Bihar during June 2008), the first intervention intended to provide remedial instruction to academically weak children in grades 3-5 during a one-month summer camp, which was conducted in school buildings by government school teachers. In reality, however, children in grades 1-5 were reached by the camp, and therefore we use this sample in our analysis. Pratham provided materials and orientation for government school teachers, and also trained volunteers 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 first model (implemented only in Bihar) involved the distribution of Pratham materials with no additional training or support (refereed to hereafter as the **Materials (or M)** treatment). The second variant of the intervention included materials, as well as training of teachers in Pratham methodology and monitoring by Pratham staff (referred to as the **Teachers and materials (TM)** treatment)**.** The third and most intensive intervention included **Materials, training, and volunteer support** (**TMV)** treatment. In addition to the materials and training components of the other interventions, in the TMV treatment villages, volunteers were recruited to provide additional support, especially in working with children who needed the most help with basic reading and arithmetic. 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. School teachers were also consulted on which children should be sent to the volunteers for extra support. This was thus the closest replication of the standard Pratham Model that had been previously evaluated in Uttar Pradesh (Banerjee et al., 2010). In Uttarakhand, volunteers supported teachers in their work within the school day, and did not work outside of school. Teachers were not only trained on how to improve basic reading and arithmetic, but were also trained in using the TaRL methodology for basic skills.

2.2.3 Evaluation Design

In the Bihar sample, schools in about 120 villages were randomly assigned to receive summer camps, while 40 served as a control group and were not offered a summer camp. 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 control group from the summer period remained the control group during the school-year evaluation. In Uttarakhand, none of the villages received summer camps, and they were randomly assigned to the control group, 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. Overall approximately 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. These surveys consisted of interviews of headmasters, teachers, and observations 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, enumerators undertook unannounced visits to all schools to observe school conditions and track program implementation. These surveys involved collecting 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, 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. In Uttarakhand, the survey team was worse, and evidence of forgery of data by surveyors was discovered. The initial baseline had to be discarded, and subsequent surveys were self-managed. In order to get baseline information before the intervention started, a shorter baseline survey was undertaken that only involved the ASER test of children.

Research 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 research staff involving semi-structured interviews with teachers, focus groups of teachers in resource centers, and focus groups of mothers in a small sample of schools and villages in both Bihar and Uttarakhand (Sharma and Deshpande, 2010).

2.3 Haryana: Teacher Training with Supervisory Support and Dedicated School Hour

As shown in the results below, of the various interventions tested in 2008-2010 in Bihar and Uttarakhand, those involving schools during their regular functioning did not lead to significant score gains, while those relying on a dedicated cadre during a dedicated period (i.e. volunteers outside of schools and government school teachers during summer camps) were effective. Pratham continued to explore models to integrate their pedagogy into the formal educational system, while confronting the inertia of traditional teaching practices. In the continuing evolution of the Read India model, Pratham's next step was to enable a cadre of government officials who could train, mentor, monitor and provide on-site support to teachers as they used the TaRL model in their schools, and more effort was devoted to signal that the program was not optional, but an integral part of the school routine, slotted in a specific time-period. In the 2012-13 school year, Read India's TaRL model was tested in schools in the state of Haryana.

2.3.1 Context

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% of students in grades 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 among Indian states (ASER Centre, 2011a).

Two districts in Haryana were chosen for the evaluation: Mahendragarh and Kurukshetra. Lying at opposite ends of the state, these districts were selected by the state government, Pratham, and the researchers because of their different economic profiles. While learning of primary schoolchildren in each district are similar to the statewide average (ASER Centre, 2011b), Kurukshetra is more economically developed than Mahendragarh. For example, 96% of households in Kurukshetra are electrified, compared with 80% in Mahendragarh (Census of India, 2011).

2.3.2 Program

There were several new features to the Read India model that were implemented as part of the Haryana evaluation. First, all efforts were made to emphasize that the program was fully supported and implemented by the Government of Haryana, rather than an external entity. To make this evident, one important innovation was the creation of a system of academic leaders within the government that could guide and supervise teachers as they implemented the Pratham methodology. In the state of Haryana, the existing school monitoring system included field-level supervisors and monitors, known as Associate Block Resource Coordinators (ABRCs). Although the administration had provided general guidelines on the roles and responsibilities of the ABRCs, the government typically did not provide specific training on how to operationalize these responsibilities in schools. ABRCs had also been used generally as "couriers" collecting information regarding various programs in the schools and delivering letters. 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 educational outcomes or impact. As part of the interventions, Pratham gave ABRCs four days of training and field practice. ABRCs were then

placed in groups of three in actual schools for a period of 15-20 days to "test" the Pratham methodology of grouping by level and of providing level-appropriate instruction. Once the "practice" period was over, ABRCs in turn trained the teachers that were in their jurisdiction. Pratham staff assisted and supported the ABRCs in this process.

The second important feature is that the program was implemented during a dedicated hour during the school day. Beginning in the 2011-12 school year, the Government of Haryana 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 Hindi classes using the Pratham curriculum. This was done merging and re-organizing across grades 3, 4, and 5. In the control schools, the school day was just longer and the extra time was used to conduct classes using their standard pedagogy. Reserving the same specific hour for restructuring the classroom across all schools simultaneously sent a signal that the intervention was government-mandated, broke the status quo inertia of routinely following the curriculum, made it easier to observe compliance, and made it practically easier for teachers to teach at the child's level.

At the start of the intervention, which coincided with the beginning of the school year, the government schoolteachers, guided by ABRCs, administered a brief oral assessment of each student's reading ability in Hindi. This assessment was used to group children by level for the "special period" during the school day. During the extra hour, in TaRL schools, all children in grades 3-5 were reassigned to ability-based groups and physically moved from their grade-based classrooms to classrooms based on levels as determined by the baseline assessment. Once classes were restructured into these level-based groups, teachers were allocated to the groups for instruction. Classroom reorganization of this type had been part of the teacher training in Bihar and Uttarakhand, but it received more emphasis in the Haryana training. Also, in Haryana, teachers were trained to re-organize classrooms across grades (teach by level, not by age), while in Bihar

and Uttarakhand the class groups were to be done within each class.

2.3.3 Evaluation Design

Four blocks, two in each of the districts, were randomly selected as the intervention sites. Across these four blocks, a total of 400 schools were randomly drawn from a list of all 467 government schools located in the four blocks.

The study of the TaRL model was conducted alongside an evaluation of another government program: Continuous and Comprehensive Evaluation (CCE), which involved training government teachers to regularly assess and provide highly detailed feedback on student performance across both curricular and extracurricular activities. The 400 schools were assigned in equal proportions to either a control group, TaRL only, CCE only, and TaRL combined with CCE. We find no detectable effect of the CCE program, as reported in Berry, et. al. (2014). In the analysis that follows, we include (but do not report) dummies for assignment to the CCE program.

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 (and would therefore be in grades 2-5 during the intervention). Tests were administered to a maximum of 10 randomly selected students in each grade in each school (which had on average 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 for the evaluations in Bihar and Uttarakhand, tested 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, as well as 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 consisted of close monitoring of all training sites where teachers were taught the TaRL methodology and surprise visits to each of the 400 evaluation schools by trained J-PAL monitors between August 2012 and March 2013. During school 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 within each school for thirty minutes to collect data on teaching and evaluation practices in the classroom.

2.4 Uttar Pradesh: In-School Learning Camps

2.4.1 Context

Uttar Pradesh is the most populous state in India, with a population of 199.8 million as of the 2011 census (Census of India, 2011b). Learning levels are particularly low in the state: as of 2011, only 48 percent of children in grades 3-5 could read grade-1 level text, the second-lowest percentage of all Indian states (ASER Centre, 2011a). The interventions in Uttar Pradesh took place in the districts of Sitapur and Unnao, both in rural areas in the center of the state. These

districts were chosen because they are large enough to support survey activity and Pratham volunteer mobilization, but they did not have any previous Pratham activity. Learning levels in Unnao are near the statewide average, while they are slightly lower in Sitapur (ASER Centre, 2011b, 2012).

2.4.2 Programs

Given the level of learning in Uttar Pradesh, the weakness of the educational bureaucracy in the state, and the high level of absenteeism in school, Pratham's assessment was that there was not enough capacity in Uttar Pradesh to take over the TaRL program following the Haryana model. Therefore, they designed a version of the program that was implemented by volunteers, but could take place within schools, allowing it to reach the largest possible number of children. Guided by the experience in Uttarakhand, the key was, once again, to create a clearly delineated space and time for the program and to also avoid seeing the volunteers simply being absorbed by the schoolteachers. To create this space, the program was conducted in the form of "Learning Camps" – intensive bursts of teaching-learning activity using the Pratham methodology. During the camps, all children were grouped according to their existing level of learning achievement, and the activities and material were designed to move children to the next level on the ASER test. The camps were led by Pratham volunteers, who all but took over the school during that short period. Pratham staff also regularly monitored the camps in each school and assisted the volunteers in administering the camps.

2.4.3 Evaluation Design

Sample schools in four blocks of Sitapur and Unnao were selected and randomly divided into either a 10-day camp treatment, a 20-day camp treatment, a treatment group that received only Pratham learning materials, or a control group, with approximately 120 schools in each group. In the 10-day camp treatment group, schools received four 10-day rounds of camp, whereas for the

20-day camp treatment group, schools received two 20-day rounds. In both groups, a 10-day booster camp was held during summer vacations. Permission was obtained from the Uttar Pradesh school administration to implement these camps within school premises during school hours. On "camp" days, children from grades 3-5 were grouped according their ability level and taught Hindi and math for about 1.5 hours each by Pratham team members and assisted by trained village volunteers. The volunteers were closely supervised by Pratham staff. In the materials-only intervention (M), schools were provided with Pratham learning materials to be used by the teachers and distributed to the students. No other form of academic support was provided by volunteers or Pratham staff.

The tests were administered to all students enrolled in grades 3-5. In addition to the child assessments, a headmaster survey was conducted during baseline as well as endline. The headmaster survey included questions on their perceptions regarding student's learning outcomes, attitude towards low performing students and activities carried out in schools to help low performers. As with the other evaluations, a systematic process monitoring survey was set-up to collect data on attendance, evidence of learning by "grouping," activities during "camp" sessions, teaching practices of volunteers, involvement of school teachers, and the school teachers' perceptions of "camp" activities.

3 Results

3.1 Descriptive Statistics

In each location, the analysis sample consists of the group of children targeted by the program. Because the programs targeted slightly different groups of children, the samples differ across locations. The Bihar summer camps and Bihar and Uttarakhad school year program included teachers and children from grades 1-5 We use the sample of children who were between the grades 1 and 5 in both years of the programs. The Haryana and Uttar Pradesh interventions

targeted children in grades 3-5 in school.

Table 2 displays basic summary statistics of key baseline variables across treatment and control groups in the four intervention areas. There are few differences across the intervention areas in terms of age, grade, and test scores. Average Hindi test scores range from about 1.1 in the Bihar sample (just above the letter level), to 1.6 in Haryana (between the letter and word levels). Math test scores average below 1 (one-digit number recognition) in all four areas.

Table 2 also evaluates the balance of these baseline variables across treatment and control groups. The Uttarakhand, Haryana, and Uttar Pradesh samples show no significant differences across these variables (at the 10% level). On the other hand, in the Bihar sample, for 3 out of 6 variables, the F-test rejects equivalence across treatment and control groups at the 5% level.

Appendix Table 1 displays attrition patterns across treatment and control groups in program locations. Attrition ranged from about 2.1 percent in Uttar Pradesh, to as high as 20 percent for the M and TM intervention groups in Bihar. As shown in the last row of each panel, attrition is not significantly related to treatment group assignment in any location.

3. 2 Empirical Specifications

Our results are based on the following specification:

$$y_{ist} = \alpha + T_s\beta + \delta y_{ist_0} + X_{is}\vartheta + \varepsilon_{ist} \qquad (1)$$

where y_{ist} represents the normalized test score for student i in village/school s at endline t (normalized by subtracting the endline mean in the control group, and dividing by the 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 randomization (village for Bihar and Uttarakhand, school

campus for Haryana, and school for Uttar Pradesh). 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.

3.3 Main Results

Table 3 displays the main results of each experiment on language and math scores. As shown in Panel A, the Bihar summer camp intervention resulted in an increase in language scores of 0.09 standard deviations (significant at the 5 percent level), and an increase in math scores of 0.07 standard deviations (significant at the 10 percent level). We note that these results are observed even though exposure to the summer camps was limited: only about 23 percent of children in treatment villages attended the camps.

Panel B displays the impacts of the three Bihar school-year interventions. Among the three treatment groups, only the combined TMV intervention resulted in a statistically significant increase in test scores, with a 0.13 standard deviation increase in language and a 0.11 standard deviation increase in math. The M and TM interventions did not have statistically significant impacts on either math or language. Since the M and TM interventions seem to have had no effect, the difference in the TMV intervention is likely due to the volunteers only, working outside of school. These results therefore replicate the findings of the original Uttar Pradesh study detailed in Banerjee et al. (2010), which, as discussed above, also found that an intervention using volunteers to teach camps outside of schools led to significant gains in reading skills.

Panel C displays the results for the two Uttarakhand treatments. Neither the TM nor the TMV interventions resulted in statistically significant increases in language or math scores

As shown in Panel D, the Haryana TaRL intervention resulted in a 0.15-standard deviation

increase in language test scores (significant at the 1 percent level). The program did not focus on mathematics at all, and correspondingly we find that it had no significant impact on math test scores.

Finally, Panel E of Table 3 displays the impacts of the interventions in Uttar Pradesh. The 10-day camp intervention increased both language and math test scores by 0.7 standard deviations, while the 20-day camp intervention increased language and math test scores by 0.6 standard deviations. These results are all significant at the 1 percent level. The M intervention also had a small, but significant, impact on math scores, with an increase of 0.04 standard deviations (significant at the 5 percent level).

3.4 Heterogeneity by Pretest Score

To examine whether the interventions were most effective for the highest- or lowest-performing children at baseline, we disaggregate impacts by baseline ASER level. Table 4 displays treatment effects (in levels moved) for each baseline level in language and in math. For language, the four classification levels are Beginner (child cannot recognize letters), Letter (child recognizes letters), Word or Paragraph (child recognizes words or can read a paragraph with grade 1 level text), or Story (child can read a story with grade 2 level text). Across the interventions with statistically significant impacts on aggregate language scores—Bihar Summer camp, Bihar TMV, Haryana TaRL, and Uttar Pradesh 10-day and 20-day learning camps—there are no consistent patterns of monotonically increasing or decreasing impacts by baseline language level. Instead, the largest impacts occurred for children at the letter level at baseline, while those who could not recognize letters (beginner level) typically experience the second largest impact.

For math, the baseline levels are: Beginner (cannot identify single-digit numbers), Single or Double Digit Number (one-digit and two-digit number recognition), Subtraction, and Division. Again, among the treatment groups that show statistically significant impacts on aggregate math

scores, there are no patterns of monotonically increasing or decreasing effects across baseline math levels. For Bihar TMV, effects are largest among those scoring lowest at baseline, while the Uttar Pradesh interventions show the largest impacts among those who could recognize numbers at baseline.

3.5 Process Data

To shed light on the intensity of implementation of each program, we turn to process data collected by enumerators. Process data were collected for the school year interventions in all four project sites. In Bihar and Uttarakhand, each school was visited once during the year, while in Haryana and Uttar Pradesh each school was visited at least twice. Enumerators 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 during each visit. In Haryana, enumerators observed a randomly-selected class for 30 minutes. In Uttar Pradesh, enumerators would spend an entire day in schools during "camp" days to allow them to observe both language and math classes.

We focus our analysis on teacher attendance and classroom observations among the interventions with schoolteachers in Bihar, Uttarakhand, and Haryana. (Monitoring data from the Uttar Pradesh intervention are still being processed.) The monitoring surveys in each state contained different questions about classroom practices; however, here 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, 67% of the teachers in the TM and TMV groups had been trained, while only 1-4% were trained in the control and M groups. In Uttarakhand, however, teacher attendance in training was much 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 96% 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 5 also displays the percentage of observed classrooms which were seen using Pratham materials. In both Bihar and Haryana treatment groups, a substantial number of teachers used Pratham materials. In Bihar, use was highest in the TM and TMV treatments, where 56-64% 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. Rates of usage were much lower among Uttarakhand treatment groups: Only 25.7% and 33.8% of classrooms used Pratham materials in the TM and TMV groups, respectively.

Finally, we examine the arrangement of students in each school. As described above, the Pratham methodology calls for grouping of students by learning levels, rather than by grade in school. Table 5 further displays the observed classroom arrangements in the three states. Here the differences between Bihar/Uttarakhand and Haryana are the starkest: in very few cases (10 percent or less) were students grouped by learning levels in Bihar and Uttarakhand, 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 TaRL 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 and Haryana, and substantial use of Pratham materials in both these states. In Uttarakhand, however, the number of trained teachers and the percentage of classrooms using Pratham materials were much lower. 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 Discussion

There results confirm that the core Pratham pedagogy, when implemented, has significant impact on learning outcomes: this was found in Uttar Pradesh (Banerjee et al., 2010), Bihar, Haryana, and Uttar Pradesh again. This method can be effectively implemented by village-level volunteers without formal teacher training, and by existing government teachers, after they have received a relatively short training on how to do this. The key challenge to mainstreaming the program in government schools was the tendency to revert back to the traditional curriculum and school organization, which occurred in some locations when the program was implemented in schools and during the normal school year.

In the original implementation in Bihar and Uttarakhand, classrooms were typically not re-organized by level of learning at any point during the class, and no specific time was assigned to the activities. 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 activities other than ensuring effective teaching, let alone organizational arrangements in treatment schools (Sharma and Deshpande, 2010).

The Haryana program addressed the core problems identified in the Bihar and Uttarakhand implementation. The TaRL treatment took place during the regular school year, with a dedicated hour specified in the school schedule to implement the program. Additionally, during the training of ABRCs, the importance of sorting children by their actual learning levels was explicitly emphasized. This was backed by a concerted effort by Pratham and research staff to give the ABRCs specific functions in implementation. Pratham's staff ensured ABRCs did 15-20 days of

practice classes before teachers were trained, with the program designed to be led by ABRCs. In addition, research 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. The process monitoring data suggest 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 or lack 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 during the dedicated hour, for which the additional effort on training and monitoring of ABRCs, as well as the clear demarcation of a specific time, appear to be key.

In parallel, Pratham has developed a model to work with volunteers in school, and with all the children. The Uttarakhand TMV program failed for very much the same reason as the TM program: Pratham volunteers were absorbed as general-purpose teacher assistants and were not given any space to conduct the TaRL activities, e.g. grouping children by level and focusing on their current achievement. However, in the original implementation in Jaunpur, Uttar Pradesh (2005-6), or in Bihar, few children were reached because they had to be coaxed in coming after school.

In Uttar Pradesh (2013-4), volunteers worked in school, with all children, but during a

dedicated time period (40 days). During that time, the schools were essentially taken over by the volunteers. Pratham volunteers and staff led the implementation, with very little involvement of government school teachers. Process monitoring results from Uttar Pradesh indicate that around 60% of school teachers never got involved in the camps even though they happened in school during school hours. Even among the teachers who did get involved, their involvement was restricted to taking student attendance and maintaining discipline. From our conversations with teachers, it seemed that while they found the method effective and materials interesting, they did not think that adopting them was a part of their "core" responsibility.

It is quite remarkable that 40 days of active teaching lead to such large learning gains: The Uttar Pradesh gains (0.7 standard deviation in both language and math) are enormous, and far and away the largest of all the interventions we tested. Figures 1 and 2 summarize visually the results in Haryana and Uttar Pradesh. The treatment effect is so large that by endline, treated children entirely catch up with the Haryana control children, and almost reach the level of the treated children in Haryana (in Uttar Pradesh, 48% of the treated children can read at the grade 2 level at endline; in Haryana, 47.5% of the control children can, and 53% of the treatment children), despite a huge gap (20 percentage point difference) at the baseline. This reflects in part the abysmal performance of the school system in Uttar Pradesh, where very little is happening in control group schools: the number of students who cannot recognize any letter between baseline and endline in the control group fell from 34% to 24% in Uttar Pradesh, while it fell from 27% to 8% in Haryana. The number of students who can read at grade 2 level increased from 14% to 24% in Uttar Pradesh, compared with 34% to 47% in Haryana. But the fact that the children actually reach the Haryana level in Uttar Pradesh also demonstrates the relative ease with which apparently daunting learning gaps can be closed. Given the relatively low cost of working with volunteers, it means that this model is as “legitimate” a scale-up model as the one that involves teachers. Since teachers are

happy to let volunteers work (but hard to convince to change their practices), volunteers seem to be doing a remarkable job, and recruiting volunteers is not difficult, working with volunteers, which appears to be extremely effective at least in environment with low starting level of learning, may be more sustainable.

5 Conclusion

Learning levels have been "low and stuck" in India for at least a decade, despite large gains in enrollment, improvement in school facilities, and massive exit to private schools. The experiments reported here show that this appears to be largely a self-inflicted problem: in 50 days of focused teaching by lightly trained volunteers, students can catch up from close to the lowest achievement levels in India to the level of learning of the third highest achieving state in the country (Haryana). The core pedagogical principle would seem to be basic common sense: group children by level of achievement, and target activities to the level they are at.

And yet, the core difficulty that these evaluations have outlined is the resistance of both parents and the school bureaucracy to implement this approach. When the program is implemented outside of school, take up is low, and the resulting impacts, while high per child treated, is not as high as it could be on average. When the program is implemented in school, either by school teachers or by volunteer, it tends to be swallowed in business as usual: the clearest symptom being that classrooms are not re-organized by learning level.

In this paper, we present two models that were evolved over several years, and successfully take on this implementation challenge: In Haryana, teachers lead the effort, supported from within the hierarchy, and with specific instructions to implement the activities and re-organize the classroom for one hour per day. In Uttar Pradesh, Pratham volunteers, supported by the Pratham hierarchy, lead the activities during specific periods which are entirely devoted to this program. As both these models led to significant learning gains, Pratham is now actively promoting these two blueprints, both of which can now be considered for scale-up in other states.

6 References

Andrabi, T., Das, J., Khwaja, A. I., Vishwanath, T., and T. Zajonc. (2007). Learning and Educational Achievements in Punjab schools (LEAPS): Insights to Inform the Education Policy Debate. *World Bank, Washington, DC.*

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

ASER Centre. (2009). *Annual Status of Education Report (Rural) 2008.* Mumbai: Pratham Resource Center.

ASER Centre. (2011a). *Annual Status of Education Report (Rural) 2011*. New Delhi: ASER Centre.

ASER Centre. (2011b). *ASER District Performance Tables 2011*. Retrieved on 24 July 2016 from: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER_2011/DPT_2011/2011districtpage.pdf

ASER Centre. (2012). *ASER District Performance Tables 2012*. Retrieved on 22 August 2016 from: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER_2012/DPT_2012/aser2012districttablesapril42013-allstates.pdf

ASER Centre. (2014). *Annual Status of Education Report 2014 – National Findings.* New Delhi: ASER Centre.

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., and S. Khemani. (2010). Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy*, 2(1):1–30.

Banerjee, A.V., Cole, S., Duflo, E., and L. Linden. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 122(3):1235– 1264.

Census of India (2011a). Houselisting and Housing Census Data Highlights. Office of the Registrar General and Census Commissioner, India.

Census of India (2011b). Primary Census Abstract – Data Highlights. Retrieved on 3 August 2016 from: http://www.censusindia.gov.in/2011census/PCA/PCA_Highlights/pca_highlights_file/India/Chapter-1.pdf

Duflo, E., Dupas, P., and M. Kremer. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101:1739–1744.

International Growth Centre (IGC). (2014). 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.

Mukherji, S., Berry, J., Duflo, E. and M. Shotland. (2014). A Wide-Angle View of Learning: Evaluation of the CCE and LEP programmes in Haryana. *Mimeo, J-PAL*.

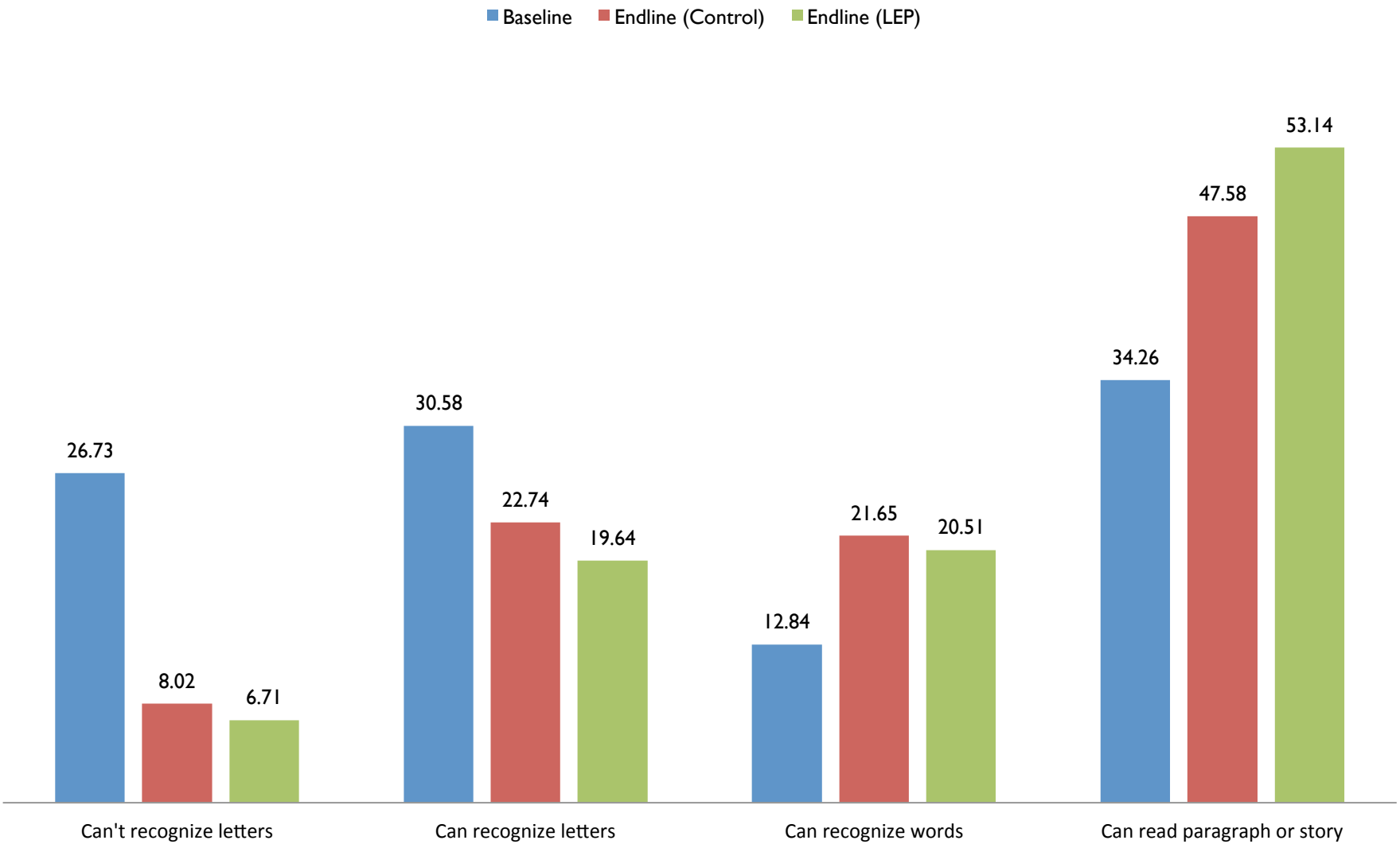
National Education Assessment (NEA). (2008). *2007 National Education Assessment*. Accra, Ghana: Ministry of Education, Science, and Sports.

Reserve Bank of India (RBI). (2011). *Handbook of Statistics on Indian Economy 2010-2011*. New Delhi: RBI.

Sharma, P. and A. Deshpande. Teachers' perception of primary education and mothers' aspirations for their children—a qualitative study in Bihar and Uttarakhand. *Mimeo, J-PAL*, 2010.

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

Distribution of Student Competency in Hindi
Baseline and Endline (By Treatment status, Haryana LEP)

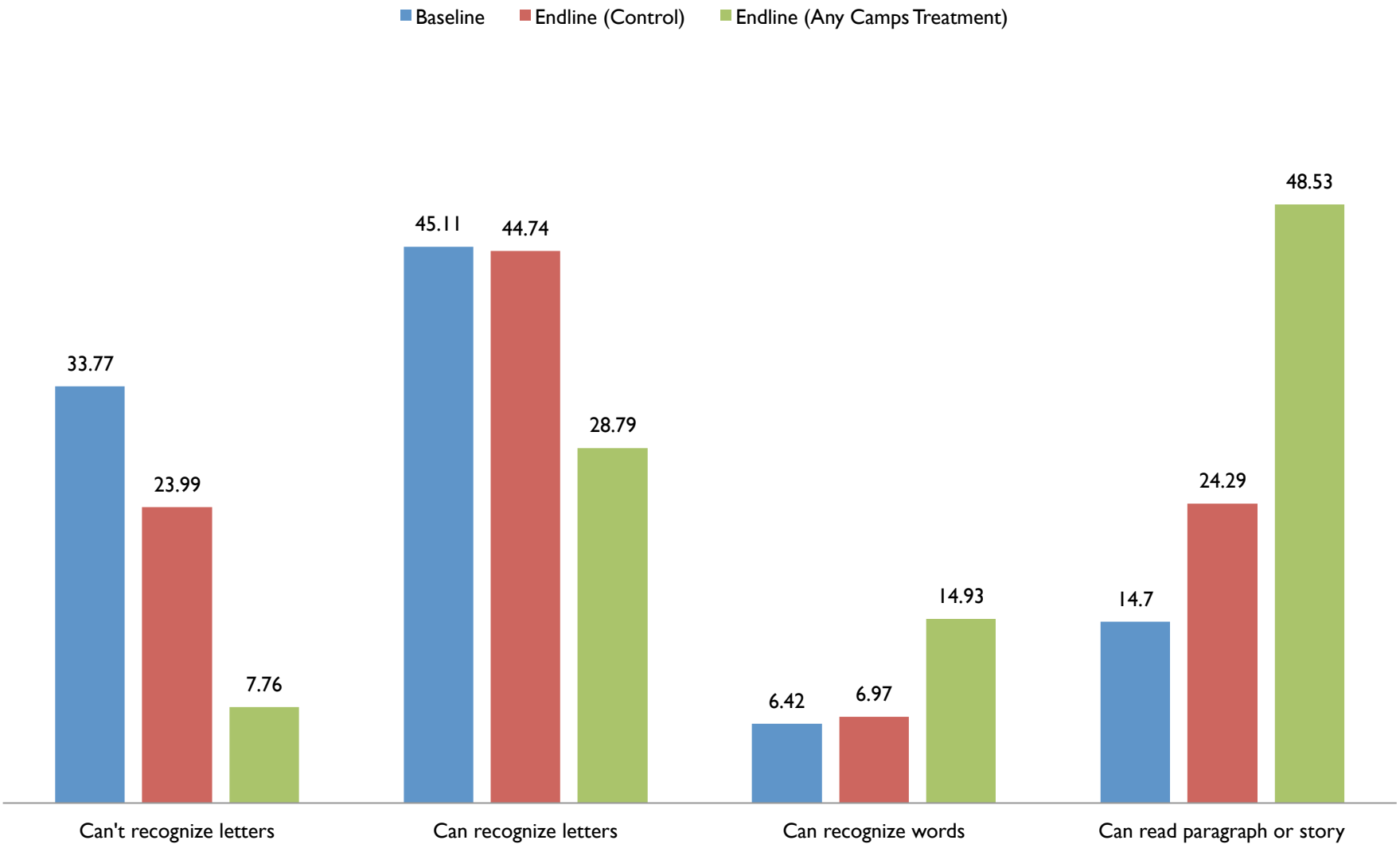


ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Pratham
Every child in school and learning well

IFMR
INSTITUTE FOR FINANCIAL MANAGEMENT AND RESEARCH

Distribution of Student Competency in Hindi
Baseline and Endline (By Treatment status, UP Learning Camps)



ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Pratham
Every child in school and learning well

IFMR
INSTITUTE FOR FINANCIAL MANAGEMENT AND RESEARCH

Table 1: Sample Composition

State	Number of Schools	Number of Students	Grades	Baseline ASER Scores	
				Language	Math
Bihar	158	6493	1-5	1.151	1.168
				(1.422)	(1.215)
Uttarakand	114	3769	1-5	1.372	1.240
				(1.439)	(1.046)
Haryana	400	11966	2-5	1.623	1.575
				(1.418)	(1.036)
Uttar Pradesh	484	17266	3-5	1.048	1.231
				(1.198)	(1.013)

Bihar and Uttarakand interventions were randomized by village (N=160 in Bihar and N=120 in Uttarakand); Haryana interventions randomized by school campus (N=400); and Uttar Pradesh interventions were randomized by school (N=484). Grades indicate the grades of students during the interventions. Baseline surveys in Haryana took place among children in grades 1-4 the year before the interventions took place.

Table 2: Balance Check of Randomization

	Female (%)	Age (years)	Grade at baseline	Out of school	Baseline oral Hindi test score (out of 4)	Baseline oral math test score (out of 4)	Number of Students	Number of Villages or Schools*
A. Bihar								
Control	53.29	8.49	2.11	0.03	1.12	1.16	1199	39
	(49.91)	(2.03)	(1.17)	(0.16)	(1.38)	(1.19)		
M	52.02	8.41	1.98	0.03	1.04	1.03	1140	37
	(49.98)	(2.01)	(1.12)	(0.17)	(1.37)	(1.15)		
TM	56.00	8.56	2.16	0.02	1.14	1.16	1066	37
	(49.66)	(2.04)	(1.24)	(0.15)	(1.42)	(1.20)		
TMV	53.21	8.55	2.16	0.03	1.30	1.32	1137	37
	(49.92)	(1.95)	(1.20)	(0.16)	(1.51)	(1.30)		
P-value (F-Test)	0.209	0.425	0.001	0.800	0.038	0.004		
P-value (t-test) Summer camps treatment	0.799	0.814	0.883	0.827	0.484	0.880		
B. Uttarakand								
Control	44.53	8.06	2.51	0.02	1.37	1.26	1188	35
	(49.72)	(1.99)	(1.21)	(0.12)	(1.46)	(1.07)		
TM	45.82	8.07	2.44	0.00	1.37	1.23	1172	36
	(49.85)	(1.95)	(1.26)	(0.07)	(1.43)	(1.01)		
TMV	45.84	8.03	2.45	0.01	1.37	1.23	1213	39
	(49.85)	(1.94)	(1.24)	(0.11)	(1.44)	(1.05)		
P-value (F-test)	0.733	0.921	0.440	0.118	0.996	0.930		
C. Haryana								
Control	50.70	9.04	2.57	0.00	1.63	1.60	2998	88
	(50.00)	(1.57)	(1.11)	(0.00)	(1.41)	(1.06)		
TaRL	50.87	9.06	2.55	0.00	1.63	1.57	6332	183
	(50.00)	(1.64)	(1.11)	(0.00)	(1.41)	(1.04)		
P-value (t-test)	0.995	0.409	0.949	.	0.742	0.864		
D. Uttar Pradesh								
Control	51.75	10.34	3.95	.	1.05	1.24	4473	123
	(49.97)	(1.24)	(0.82)	.	(1.20)	(1.03)		
M	51.70	10.25	3.98	.	1.05	1.25	4284	119
	(49.98)	(1.27)	(0.81)	.	(1.22)	(1.03)		
10-day Camp	50.73	10.26	3.94	.	1.03	1.21	4532	122
	(50.00)	(1.32)	(0.82)	.	(1.18)	(0.99)		
20-day Camp	50.72	10.31	3.97	.	1.06	1.23	4310	120
	(50.00)	(1.32)	(0.82)	.	(1.20)	(1.01)		
P-value (F-Test)	0.827	0.254	0.110	.	0.973	0.915		

Note: Standard deviations in parentheses.

*Number of villages or schools as per level of randomation (school level for Haryana and Uttar Pradesh; village level for Bihar and Uttarakhand)

M = Materials, TM = Teachers and materials, TMV = Materials, training and volunteer support, TaRL = Teaching at the right level

Table 3: Language and Math Results

	Language	Math
A. Bihar -- Summer Camp		
Treatment	0.0867**	0.0742*
	(0.0417)	(0.0440)
Observations	2839	2838
B. Bihar -- School Year		
M	0.0168	0.0405
	(0.0392)	(0.0406)
TM	0.0426	0.0145
	(0.0384)	(0.0389)
TMV	0.125***	0.105***
	(0.0350)	(0.0366)
Observations	6490	6490
C. Uttarakhand		
TM	0.0636	0.0591
	(0.0410)	(0.0451)
TMV	0.0119	0.0252
	(0.0312)	(0.0441)
Observations	3763	3762
D. Haryana		
TaRL	0.154***	-0.00611
	(0.0173)	(0.0170)
Observations	11963	11962
E. Uttar Pradesh		
M	0.0336	0.0449**
	(0.0219)	(0.0228)
10-Day Camp	0.701***	0.694***
	(0.0224)	(0.0242)
20-Day Camp	0.609***	0.620***
	(0.0229)	(0.0243)
Observations	17254	17265

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.
M = Materials, TM = Teachers and materials, TMV = Materials, training and volunteer support, TaRL = Teaching at the right level

Table 4: Levels Moved by Baseline Level

	Language					Math				
	All Levels	Nothing	Letter	Word or Paragraph	Story	All Levels	Nothing	Single or Double Digit Number	Subtraction	Division
A. Bihar -- School Year										
M	0.0181	0.0150	0.0584	-0.0361	0.0325	0.0303	-0.00322	0.0412	0.0947	-0.0462
	(0.0410)	(0.0546)	(0.0772)	(0.0705)	(0.0716)	(0.0359)	(0.0456)	(0.0432)	(0.107)	(0.111)
TM	0.0641	0.0691	0.0977	0.0197	-0.0223	0.00303	-0.0279	0.0330	0.0438	0.00484
	(0.0429)	(0.0541)	(0.0873)	(0.0747)	(0.0961)	(0.0352)	(0.0426)	(0.0521)	(0.116)	(0.0776)
TMV	0.132***	0.172***	0.182**	0.0851	0.0175	0.0486	0.0642	0.0225	0.0355	0.0762
	(0.0418)	(0.0639)	(0.0800)	(0.0616)	(0.0679)	(0.0353)	(0.0488)	(0.0455)	(0.0905)	(0.0825)
Observations	4500	2174	1006	764	556	4483	1706	1983	546	248
B. Uttarakhand										
TM	0.0671	0.0717	0.0799	0.0576	-0.0192	0.0360	-0.0103	0.0460	0.0644	0.0240
	(0.0453)	(0.0852)	(0.0634)	(0.0563)	(0.0197)	(0.0361)	(0.0388)	(0.0493)	(0.0971)	(0.104)
TMV	-0.0157	-0.0743	0.0263	0.0420	-0.0298	0.0355	-0.0144	0.0373	-0.0237	0.00132
	(0.0336)	(0.0726)	(0.0523)	(0.0507)	(0.0253)	(0.0351)	(0.0350)	(0.0432)	(0.0949)	(0.0915)
	0.569	0.700	0.741	0.531	-0.0330	0.445	0.831	0.358	0.291	-0.185
Observations	3551	1285	1100	631	535	3551	899	2221	288	143
C. Haryana										
TaRL	0.138***	0.167***	0.169***	0.132***	0.0288	-0.0141	0.0273	-0.0170	-0.0871*	0.00900
	(0.0186)	(0.0394)	(0.0272)	(0.0246)	(0.0195)	(0.0177)	(0.0269)	(0.0195)	(0.0444)	(0.0458)
Observations	11876	3162	3627	3168	1919	11870	1232	8497	1371	770
D. Uttar Pradesh										
M	0.0365	0.000163	0.0792***	0.0948**	-0.0628	0.0457**	0.0631**	0.0418*	0.0558	0.0707
	(0.0244)	(0.0288)	(0.0306)	(0.0469)	(0.0384)	(0.0192)	(0.0317)	(0.0214)	(0.0743)	(0.0630)
10-Day Camp	0.680***	0.771***	0.792***	0.446***	0.0480	0.536***	0.369***	0.605***	0.422***	0.226***
	(0.0249)	(0.0356)	(0.0310)	(0.0414)	(0.0298)	(0.0255)	(0.0341)	(0.0308)	(0.0671)	(0.0575)
20-Day Camp	0.582***	0.648***	0.711***	0.370***	0.0431	0.482***	0.327***	0.544***	0.466***	0.204***
	(0.0260)	(0.0366)	(0.0322)	(0.0463)	(0.0262)	(0.0273)	(0.0351)	(0.0320)	(0.0686)	(0.0608)
Observations	17182	6686	6859	2283	1354	17215	3057	12292	677	1189

Standard errors in parentheses (clustered at level of randomization). Regressions control for baseline test scores, as well as gender, age, and standard
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
** Significant at the 5 percent level. *** Significant at the 1 percent level. M = Materials, TM = Teachers and materials, TMV = Materials, training and volunteer support, TaRL = Teaching at the right level. Observations that are missing baseline values are not included.

Table 5: Selected Process Results

	Percent of Observations		
	Teachers Trained	Pratham materials used	Classes grouped by ability
A. Bihar -- School Year			
Control	1.0	0.0	0.0
	(63)	(110)	(99)
M	3.9	31.4	0.9
	(64)	(121)	(112)
TM	67.3	57.5	4.1
	(66)	(134)	(121)
TMV	67.1	64.3	0.0
	(68)	(126)	(116)
B. Uttarakhand			
Control	15.2	2.6	11.4
	(41)	(78)	(70)
TM	27.5	25.7	10.1
	(40)	(74)	(69)
TMV	44.9	33.8	5.9
	(39)	(74)	(68)
C. Haryana			
Control	0.2	1.0	0.0
	(198)	(198)	(193)
TaRL (During specified time)	96.0	73.8	91.7
	(84)	(84)	(84)
TaRL (Other times)	93.7	1.9	2.0
	(106)	(106)	(100)
D. Uttar Pradesh Camps			
Control		0.0	
		(189)	
M		30.5	
		(187)	
10-Day Camp	91.0	88.4	81.0
	(76)	(311)	(253)
20-Day Camp	87.5	81.4	82.7
	(83)	(312)	(243)

Notes: 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. Number of observations in parentheses. M = Materials, TM = Teachers and materials, TMV = Materials, training, and volunteer support, TaRL = Teaching at the right level. For Haryana, note that treatment schools were either visited during specified TaRL time or other times (but not both).

Appendix Table 1: Attrition

	Attrition (%)	Observations
A. Bihar (School year, endline)		
Control	0.154	251
	(0.361)	
TMV	0.166	260
	(0.373)	
TM	0.203	314
	(0.402)	
M	0.201	326
	(0.401)	
P-value (F-Test)	0.315	
B. Uttarakand		
Control	0.085	134
	(0.280)	
TMV	0.073	104
	(0.259)	
TM	0.094	133
	(0.292)	
P-value (F-test)	0.359	
C. Haryana		
Control	0.052	157
	(0.223)	
TaRL	0.048	303
	(0.214)	
P-value (t-test)	0.883	
D. Uttar Pradesh		
Control	0.022	100
	(0.148)	
M	0.019	80
	(0.135)	
10-Day Camp	0.018	82
	(0.133)	
20-Day Camp	0.023	100
	(0.151)	
P-value (F-test)	0.336	

Note: Standard deviations in parentheses. M = Materials, TM = Teachers and materials, TMV = Materials, training, and volunteer support, TaRL = Teaching at the right level

Appendix Table 2: Summary of Experiments and Results

	Bihar	**Uttarakhand**	**Haryana**	**Uttar Pradesh**
Time Period:	2008-2010	2008-2010	2012-2013	2013-2014
Experiments:	**(1) Summer Camp (June 2008)**: Remedial instruction for academically weak 1st through 5th grade students. Pratham provided materials and training of government teachers. Trained volunteers supported teachers in the classroom. The following three interventions were conducted during the shool year for children in grades 1-5: **(2) Materials (M)**: Pratham materials were distributed without any additional support **(3): Teachers and Materials (TM)**: Materials and training of teachers in Pratham methodology and monitoring by Pratham staff. **(4) Training, Materials, and Volunteering Support (TMV):** Materials and training components plus additional support provided by village volunteers. In Bihar, the volunteers worked outside of school hours.	**(1) TM:** Same as Bihar **(2) TMV:** Same as Bihar, except volunteers supported teachers in school during the day	**(1) Teaching at the Right Level (TaRL)**: Adaptation of the Pratham model - Pratham trained teachers, Pratham materials were used in in schools, and the program was implemented during a dedicated hour each day. Schools were supervised and monitored by another set of trained government personnel. **(2) Continuous Comprehensive Evaluation (CCE):** Government program that trained government teachers to regularly assess and provide feedback on student performance. **(3) TaRL + CCE**	For both the 10-day and 20-day interventions, volunteers implemented the Pratham program in the form of intense "Learning camps." Children were grouped according to their learning levels with the appropriate materials. Pratham staff regularly monitored the camps and assisted volunteers. The interventions differed in the length of time in which they were implemented: **(1) 10-day intervention**: Group recieved four 10-day rounds of camp **(2) 20-day intervention**: Group received two 20-day rounds of camps For both (1) and (2) a 10-day booster camp was also held during summer vacations. (3) **Materials (M):** Another treatment arm received only Pratham materials without any additional support
Summary of Results:	Summer camp and TMV interventions significantly improved both language and math scores. M and TM interventions had no effect.	TM and TMV interventions had no effect on language or math scores	There was no detectable effect of the CCE program, as reported in Berry et al. (2014). The TaRL intervention resulted in a significant increase in language scores, but not improvement in math scores.	10-day and 20-day camps significantly increased both language and math scores. The M intervention had a small, but significant, impact on math scores.
Implementation Notes:	For M, TM, and TMV classrooms were never organized around initial learning levels.	Classrooms were also not organized around initial learning levels. Volunteers were used by teachers as assistants to carry out regular activities.	Classrooms were successfully reorganized.	Intense camps allowed Pratham volunteers to literally take over the school during the period of time they ran the camps.