Esther Duflo
James Berry
Shobhini Mukerji
Marc Shotland

A wide angle view of learning

Evaluation of the CCE and LEP programmes in Haryana, India

February 2015

Impact Evaluation Report 22

Education

3ie International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Impact Evaluations

3ie-supported impact evaluations assess the difference a development intervention has made to social and economic outcomes. 3ie is committed to funding rigorous evaluations that include a theory-based design, use the most appropriate mix of methods to capture outcomes and are useful in complex development contexts.

About this report

3ie accepted the final version of this report as partial fulfilment of requirements under grant OW4.1005 issued under Open Window 4. The content has been copyedited and formatted for publication by 3ie. Due to unavoidable constraints at the time of publication, a few of the tables or figures may be less than optimal. All of the content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its Board of Commissioners. Any errors and omissions are also the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, James Berry, at jwb295@cornell.edu.

Funding for this impact evaluation was provided by 3ie's donors, which include UKaid, the Bill & Melinda Gates Foundation, Hewlett Foundation and 12 other 3ie members that provide institutional support. A complete listing is provided on the 3ie website at http://www.3ieimpact.org/about-us/3ie-members.

Suggested citation: Duflo, E, Berry, J, Mukerji, S and Shotland, M, 2015. *A Wide Angle View of Learning: evaluation of the CCE and LEP Programs in Haryana, India, 3ie Impact Evaluation Report 22.* New Delhi: International Initiative for Impact Evaluation (3ie)

3ie Impact Evaluation Report Series executive editors: Jyotsna Puri and Beryl Leach
Managing editor: Omita Goyal
Assistant managing editors: Radhika Menon and Markus Olapade
Production manager: Pradeep Singh
Copy editor: Proteeti Banerjee
Proofreader: Mathew PJ
Cover design: John F McGill
Printer: Via Interactive
Cover photo: Global Partnership for Education and Deepa Srikantaiah / Flickr

A wide angle view of learning: evaluation of the CCE and LEP programs in Haryana, India

Esther Duflo
Massachusetts Institute of Technology

James Berry
Cornell University

Shobhini Mukerji
Abdul Latif Jameel Poverty Action Lab (J-PAL) South Asia at IFMR

Marc Shotland
J-PAL

3ie Impact Evaluation Report 22

February 2015

Acknowledgements

This evaluation was conducted in collaboration with the Government of Haryana and the NGO Pratham. We are grateful to John Firth, Harini Kannan, Kanika Dua, Shweta Rajwade and Perwinder Singh for their superb work coordinating research activities in the field and to Harris Eppsteiner for excellent research assistance. Research discussed in this publication was funded by the Government of Haryana, by CLEAR (Regional Centers for Learning on Evaluation and Results), and by the International Initiative for Impact Evaluation (3ie) through the Global Development Network (GDN). All errors are our own. The views expressed in this article are not necessarily those of 3ie, its members, CLEAR, or the Government of Haryana.

Abstract

We report the results of a randomized evaluation of two programs designed to improve student achievement in primary and upper primary schools in Haryana, India. In one program, Continuous and Comprehensive Evaluation (CCE), high-stakes exams are replaced with more frequent evaluation of students by teachers. In the other program, the Learning Enhancement Program (LEP), students are given a brief assessment of basic Hindi skills at the start of the academic year, and a portion of the school day is set aside to group and teach students according to ability level, regardless of age or grade. Four hundred primary schools were randomly assigned to one of four groups that received (1) CCE alone, (2) LEP alone, (3) CCE and LEP together, or (4) no treatment. An additional 100 upper primary schools were randomly assigned to receive either (1) CCE alone or (2) no treatment. We find that students in primary schools assigned to receive LEP perform 0.152 standard deviations better on oral tests of basic Hindi and 0.135 standard deviations better on written tests of basic Hindi than the control group. The CCE program had no significant effect on test scores for students in either primary or upper primary schools, and there was no significant effect of combining the two programs relative to LEP alone. Neither program, either alone or in combination, had a significant effect on math test scores. LEP's large effect on students' basic Hindi skills indicates that programs emphasizing teaching at the right level can play a role in improving poor learning outcomes in India.

Abstract

We report the results of a randomized evaluation of two programs designed to improve student achievement in primary and upper primary schools in Haryana, India. In one program, Continuous and Comprehensive Evaluation (CCE), high-stakes exams are replaced with more frequent evaluation of students by teachers. In the other program, the Learning Enhancement Program (LEP), students are given a brief assessment of basic Hindi skills at the start of the academic year, and a portion of the school day is set aside to group and teach students according to ability level, regardless of age or grade. Four hundred primary schools were randomly assigned to one of four groups that received (1) CCE alone, (2) LEP alone, (3) CCE and LEP together, or (4) no treatment. An additional 100 upper primary schools were randomly assigned to receive either (1) CCE alone or (2) no treatment. We find that students in primary schools assigned to receive LEP perform 0.152 standard deviations better on oral tests of basic Hindi and 0.135 standard deviations better on written tests of basic Hindi than the control group. The CCE program had no significant effect on test scores for students in either primary or upper primary schools, and there was no significant effect of combining the two programs relative to LEP alone. Neither program, either alone or in combination, had a significant effect on math test scores. LEP's large effect on students' basic Hindi skills indicates that programs emphasizing teaching at the right level can play a role in improving poor learning outcomes in India.

Contents

List of figures and tables

Abbreviations and acronyms

ABRC	Assistant Block Resource Coordinator
ACR	Annual Confidential Report
CBSE	Central Board of Secondary Education
CCE	Continuous and Comprehensive Evaluation
GoH	Government of Haryana
LEP	Learning Enhancement Program
NCERT	National Council for Education Training and Research
NIRD	National Institute of Rural Development
NUEPA	National University of Educational Planning and Administration
PS	Principal Secretary
RTE	Right to Education
SCERT	State Council for Education Training and Research
TaRL	Teaching at the Right Level

1. Introduction

Current enrollment rates at the primary school level in India are well over 95 per cent, and dropout rates do not appear to increase dramatically with age—children between the ages of 11 and 14 are only three percentage points more likely to be out of school than children between the ages of 7 and 10 (ASER Centre 2013). However, despite India's success in increasing student enrollment and retention rates, learning outcomes have not kept pace. National-level educational surveys have consistently shown that the vast majority of Indian students fail to attain grade-level competencies at the end of five years of primary schooling. A 2012 survey found that only 47 per cent of students in grade 5 could read grade 2 level texts proficiently, while only about 25 per cent of grade 5 students were able to solve questions involving division, a grade 4 level competency. Furthermore, as students progress through school, those who lag behind in early grades continue to fall further and further behind. Many still do not possess even basic skills at the end of eight years of primary education: only 76 per cent of grade 8 students can read a grade 2 level text, while only 48 per cent can complete mathematical operations at a grade 4 level[1]. One of the most important challenges to India's continued development is the need to address these deficits in crucial skills, and in particular to improve basic learning outcomes in early grades to provide a foundation for continued learning as children progress through school.

Evaluation practices in Indian schools may contribute to poor learning outcomes. Historically, student evaluations in India have focused almost exclusively on academic topics, with tools of evaluation restricted to end-of-the-year examinations (NCERT 2005). With these "high-stakes" exams as the sole source of information about their students' abilities, teachers may lack adequate information about students' educational achievement to address their individual needs. To improve this evaluation system, the 2009 Right to Education (RTE) Act eliminated these exams and introduced a system of Continuous and Comprehensive Evaluation (CCE) as a replacement. In the CCE framework, teachers are given rubrics and trained on how to frequently evaluate students along a wide variety of academic and non-academic dimensions. The theory underlying CCE is that better tracking of children allows teachers to customize their teaching based on the current learning levels of individual students. Indian states are currently in the process of designing CCE programs, which will be rolled out nationwide in the coming years. While many educationalists in India believe that the introduction of CCE will positively affect student achievement (CBSE 2009), there has been no systematic examination of such a system to support this hypothesis.

Moreover, as noted above, the vast majority of students in grades 1–5 do not possess grade-level competencies. While CCE may highlight this challenge, the program by itself is not designed to equip teachers with the tools to bring these children to grade level. One possible solution to this concern is the Learning Enhancement Program (LEP), developed by Pratham, a large non-governmental organization focusing on basic literacy and numeracy. Under LEP, students' literacy and numeracy levels are identified at the beginning of the year through a rapid oral test. Following this test, classes are restructured according to those levels—rather than grade—for a segment of the day, during which each skill group is taught using a curriculum designed to address its

[1] Ibid.

particular skill deficit(s). LEP's methodology and curriculum are based on Pratham's Read India program, which has been shown to be effective at improving basic skills when implemented both by Pratham staff and Pratham-trained volunteers in multiple contexts (Banerjee *et al.* 2010; Banerjee *et al.* 2011). To date, however, there has been no evidence that the methodology has similar effects when implemented by government teachers as within the formal schooling system.

To address these knowledge gaps, the Abdul Latif Jameel Poverty Action Lab (J-PAL) partnered with the Government of Haryana, a state in northwest India, to conduct a randomized controlled trial to evaluate the impact of both CCE and LEP on student achievement. The evaluation sought to answer the following questions for primary school students:

1. Does the CCE program improve student test scores relative to the status quo (without any specific training on the pedagogy of remedial education, or restructuring of classes)?
2. Does the LEP program (with a one-time assessment of learning at the beginning of the year, but without the continuous assessment found in CCE) improve student test scores?
3. Does a combination of the CCE and LEP improve student test scores, both relative to the status quo and relative to each program individually?
4. Do these impacts depend on class or child characteristics?

For upper primary school students, the evaluation was designed to answer only the first question.

Four hundred primary schools in the districts of Mahendragarh and Kurukshetra were assigned at random to receive one of four treatments during the 2012–13 academic year: (a) CCE alone, (b) LEP alone, (c) both programs, and (d) neither program. To evaluate the impact of CCE on older students, an additional 100 upper primary schools were also included in the study, and were randomly assigned either to implement CCE or to act as a control group. Student achievement was measured by both oral and written assessments conducted during the 2011–12 school year (baseline), and at the end of the 2012–13 school year (endline).

We find that students in CCE schools did not perform significantly better at endline than students in control schools on either oral or written tests, whether in primary schools or in upper primary schools. On the other hand, the LEP program had a large, positive and statistically significant effect on students' basic reading abilities: students in primary schools where LEP was implemented scored 0.152 standard deviations higher on oral tests of basic Hindi reading ability, and 0.135 standard deviations higher on written tests of basic Hindi than corresponding students in control schools at endline. LEP did not, however, have a significant effect on math scores. Finally, combining CCE and LEP had no significant effect on student test scores relative to the LEP program alone.

The remainder of this report is organized as follows: Section 2 provides context for the intervention and evaluation; Section 3 describes the intervention and our theory of change; Section 4 provides an overview of our evaluation design; Section 5 discusses data sources and student testing; Section 6 discusses the implementation of CCE and LEP; Section 7 presents impact results; and Section 8 makes policy recommendations.

2. Background and context

2.1 Continuous and Comprehensive Evaluation (CCE)

Enacted in 2009 by India's Parliament, the Right to Education Act eliminated terminal, high-stakes standardized exams in government schools, replacing them with frequent low-stakes assessments of student achievement (Government of India 2009). The Continuous and Comprehensive Evaluation scheme is intended to provide teachers and students with broad-based and frequent feedback on performance to allow teachers to customize their teaching based on the current learning levels of individual students. To this end, CCE prescribes a more "comprehensive" assessment of student achievement than traditional testing: it assigns scores not only on the basis of scholastic performance, but also on the basis of co-scholastic activities (such as arts, music, or athletics) and personality development as reflected in life skills, attitudes, and values. CCE's mode of assessment is also meant to be "continuous," in that teachers identify students' learning progress at regular time intervals on small portions of content (such as a single module or lesson). This regular assessment incorporates a variety of techniques, including unit tests, projects, and evaluation of class participation. It is designed to reduce the stress of preparing for major exams, while enabling teachers to closely monitor student progress and better tailor their teaching to student needs. Although mandated by the Central Government, details of the design and implementation of CCE—which covers students in grades 1 through 8—were made the responsibility of state-level education ministries.

CCE's methodology is based on previous work in education research emphasizing the importance of the "formative use of summative assessments," that is, the use of assessments that provide evidence of student achievement as feedback, which informs teachers' classroom practices and, in particular, their interactions with students. Black and Wiliam (2004, 2009) argue that using assessment in this way can play an important role in advancing student learning. Evaluations of programs focusing on continuous assessment have been conducted in developing-country contexts such as Malawi (Bolyard 2003; Kamangira 2003) and Zambia (Kapambwe 2010); none of these evaluations, however, used random assignment of subjects to treatment and comparison groups, making the attribution of outcomes to their respective interventions uncertain. Moreover, little work has been done to investigate whether the formative use of student evaluation can positively impact student achievement once operationalized in a large-scale institutional context like that of Haryana's public school system. Despite some Indian educationalists' enthusiastic endorsement of CCE's transformative potential (CBSE 2009), to date no rigorous evaluation of the program's impact on student learning has taken place.

2.2 The Learning Enhancement Program (LEP)

Designed by Pratham, India's largest education NGO, the Learning Enhancement Program (LEP) is based on the idea of "teaching at the right level" (TaRL): LEP provides tools and allocates time within the school day schedule to enable teachers to focus their teaching at each child's competency level. In the Pratham pedagogy, children are assessed using a simple oral tool to measure literacy and numeracy levels, and are then grouped according to that level rather than standard or age. Each group ("*mahal*") is

taught starting from its current competency level, and level-appropriate learning activities and materials are used. Throughout the entire process, children's progress is assessed through ongoing simple measurements of their ability to read, to write, to comprehend basic mathematics, and to perform basic arithmetic. As with Pratham's flagship "Read India" program, the LEP curriculum focuses on children in grades 3–5, since it is difficult to identify students who are falling behind in 1st and 2nd grade, when all children are at a relatively low level.

In previous studies, Pratham's TaRL methodology has been shown to be effective at improving basic literacy and numeracy in multiple contexts in India. Evaluations of TaRL in public primary schools in Bihar and Uttarakhand between 2008 and 2010, for example, showed significant improvements in student scores when the program was implemented by Pratham staff and/or Pratham-recruited volunteers (Banerjee *et al.* 2011). Recent evaluations in both Bihar (*ibid.*) and Uttar Pradesh (Banerjee *et al.* 2010) also found the methodology to have significant positive impacts when implemented by government teachers and Pratham volunteers, respectively, in camps designed to teach remedial reading skills outside of school. More generally, other programs focused on teaching students at the right level have also been shown to be effective at improving student abilities in basic skills in India (Banerjee *et al.* 2007), Kenya (Duflo *et al.* 2011), and Ghana (Duflo and Kiessel 2012).

2.3 Setting

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 per cent of students in classes 1 to 8 in Haryana could read a grade 2 level text, while the corresponding number was 38 per cent 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 three 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 the basic measures of child literacy and numeracy, Kurukshetra falls primarily below the state average, while Mahendragarh is above the state average. Of children aged 6–14, 0.5 per cent in Kurukshetra were not enrolled in school in 2010, compared with 0.1 per cent of children in Mahendragarh (ASER Centre 2011b). Kurukshetra is also more economically developed than Mahendragarh. For example, 96 per cent of households in Kurukshetra are electrified, compared with 80 per cent 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).

Table 1 summarizes selected differences between Kurukshetra, Mahendragarh, Haryana, and India as a whole.

Table 1: Selected development and education indicators

	Population	Literacy Rate, Ages 7+ (%)	% Households with Electric Lighting	% Children in Std. 3–5 who can read Level 1 text	% Children in Std. 3–5 who can perform subtraction	% Children out of school	% Children under 3 stunted	% Children under 3 underweight
Kurukshetra	964,655	76.3	95.8	65.2	53.4	0.5		
Mahendragarh	922,088	77.7	80.4	71.2	74.0	0.1		
Haryana (All Districts)	25,351,462	75.6	90.5	69.8	64.5	1.4	43.3	38.2
India	1,210,569,573	73.0	67.2	57.5	46.5	3.3	44.9	40.4

Source: ASER Centre 2011b; Census of India 2011; IIPS & Macro International 2008.

Haryana's state education system—which contains over 20,000 schools, 80,000 teachers, and two million students (NUEPA 2013)—is structured similarly to government school systems in other states in India, with administration on the state, district, and block (administrative units one level below districts) levels. As with other states, policy and curricular decisions are made at the state level; syllabus development, along with the content of teacher training and textbooks, is the responsibility of the State Council for Education Training and Research (SCERT), a state-level body. Under RTE, each state's SCERT has been given a broad mandate for setting general education policy and has been designated as the body in each state responsible for the implementation of CCE. As with the majority of Indian civil servants, upper-level education administrators in Haryana are hired through the Indian Administrative Service, following training at the national administration academy. Day-to-day school functioning is managed at the district and block levels.

Teachers in Haryana's government school system are required to have passed at least high school (although teachers with less than a Bachelor's degree are severely constrained in their advancement) and an eligibility test. Evaluation of teachers by headmasters and local-level administrators is fairly ad-hoc: although an Annual Confidential Report (ACR) is created for each teacher at the block level, anecdotal evidence suggests that the contents of these reports are highly subjective. Though block- or district-level officials may visit "problem" schools on an ad-hoc basis, there is little to no formal top-down monitoring of teacher behavior and/or classroom practices. On the whole, almost all teachers we encountered in informal meetings throughout the course of this study cited completion of the state-approved syllabus to be their primary concern in teaching.

3. Description of intervention

The CCE program in Haryana was conceptualized by the SCERT with significant inputs from senior administrative officials such as the Principal Secretary, the State Project Director and the Director of Elementary Education. The CCE guidelines were developed by SCERT while the training content was developed by two private-sector training companies. A state resource group comprising of senior department officials, SCERT staff, and private consultants reviewed the training content before finalization. Teachers were trained by two education training companies partnering with SCERT. This training emphasized the need for frequent evaluation of student performance as a means of tracking student progress in reaching grade-level competencies. Students were to be evaluated using a variety of new tools such as unit tests, projects, homework assignments, and assessment of class participation. The tools used varied across the grades, with only observation-based evaluation (as opposed to written work) being used for grades 1 and 2. The critical innovation of this CCE program was the introduction of "evaluation sheets" for recording evaluations of students. Evaluation sheets were to be completed every month (or quarterly for grades 6–8), while report cards would be created twice a year. In a significant break from the norm, descriptive remarks and alphabetical grades were to be provided instead of numerical marks. In addition to teacher training, CCE schools were also provided with materials such as manuals, evaluation sheets, and report cards in order to implement the program.

As implemented in Haryana, the LEP curriculum focused solely on basic Hindi skills, and did not address basic math skills explicitly (a more detailed explanation of Pratham's decision to implement only the Hindi curriculum of LEP in Haryana is provided in Section 6.2). Teacher training was conducted by Pratham staff. In LEP schools, 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. Based on the results of this assessment, all students in each LEP school were to be reassigned for part of the school day to classrooms based on these levels: 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 ability-specific "*mahal*s," teachers were to use the *mahal*-specific curriculum designed by Pratham to teach each *mahal*. The Pratham curriculum emphasizes student-teacher interactivity as a means of addressing each *mahal*'s specific needs.

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 LEP schools, the extra hour was used for class reorganization and teaching remedial classes using the Pratham curriculum. In other schools in our study, the extra time was used to conduct classes using their standard pedagogy (the status-quo curriculum, either with or without CCE).

3.1 ABRC monitoring

Since neither LEP nor CCE has been proven to work in schools when implemented by teachers alone, and since low levels of implementation could have jeopardized the impact of either program, J-PAL South Asia repeatedly stressed the importance of monitoring and management of these programs to GoH. Recognizing the value of

monitoring the content delivery in schools more effectively, GoH requested that J-PAL SA help revive its existing school-level program monitoring and management to enable the monitoring of CCE and LEP.
The existing scenario included an established system of school monitoring with block and district supervisors and field-level monitors, known as Associate Block Resource Coordinators (ABRCs). Informal discussions with officials at various levels of the Department of Education revealed that 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 until then 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 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.

Working with the state administration, J-PAL SA set up a monitoring and mentoring program for CCE and LEP using the ABRCs, who were trained on both programs and who worked with teachers in all schools included in the evaluation. As a part of these efforts, J-PAL SA in partnership with SCERT, district level officials and selected ABRCs piloted and created monitoring tools; trained ABRCs on monitoring, basic data analysis, and report writing; facilitated data sharing and identification of issues; and trained district and block-level officials in performance management of ABRCs. ABRCs helped ensure the implementation of both CCE and LEP in treatment schools, as well as the non-implementation of either program in control schools. They also collected data on teacher practices and school needs, and served as resources to teachers in implementing both programs and in teaching according to best practices. Each ABRC was responsible for between 10 and 15 schools.

The State Project Director mandated a monthly review meeting to discuss the progress of implementation of CCE and LEP. This meeting was held regularly during the course of the evaluation and attended by the ABRCs, block and district officials, representatives from SCERT, J-PAL SA and Pratham. The monthly review meeting provided a platform for ABRCs to discuss findings from their school visits. Many substantive issues were brought up by ABRCs not limiting to teacher absence, replenishment of program materials, means to involve parents, outdated teaching practices etc. The senior officials were responsive to these findings with action being taken on most aspects within a short time-frame.

3.2 Theory of change

Both CCE and LEP are intended to address the issue of poor student performance in government schools, as reflected primarily in students' failure to achieve grade-level competencies. The main beneficiaries of both programs are thus students in government schools, who, thanks to the intervention(s), are expected to improve their learning outcomes. The programs, however, are designed to address this need for improvement in different, yet potentially complementary, ways.

The premise of CCE is that, under the status quo, teachers, children, and parents lack feedback on students' learning levels and progress in school, and this in turn contributes to poor learning outcomes. In public schools, examinations are typically infrequent, and

therefore do not provide teachers with enough information on students' learning levels to effectively inform their pedagogy. Without this information, teachers do not teach to the appropriate level of the students, and instead lecture students based solely on the scheduled curriculum, without reference to their students' actual abilities. As a result, students who are behind in the curriculum may lose interest and fall further behind. CCE addresses the lack of information by training teachers to collect frequent information on student learning. Armed with this information, the theory suggests that teachers may be able to teach to the levels of children in their classes, rather than rigidly sticking to the scheduled curriculum. Moreover, continuous assessment may be more accurate than "one-shot" evaluations, since students may underperform on individual assessments if they are under large amounts of pressure or if they are simply having a bad day.

Key assumptions of the CCE intervention include:

1. Part of the constraint to student learning is a lack of information and/or a lack of teacher focus on individual student strengths and weaknesses;
2. Teachers will attend trainings for CCE, where they will understand the program's goals and methodology;
3. Teachers and school administrators are willing to put the program in place in their classrooms by frequently evaluating their students' performance;
4. Teachers have sufficient time and resources to be able to incorporate continuous evaluation into their regular teaching practices;
5. Teachers will be able to effectively use the information they gain through these evaluations to identify low-performing students, and will devise ways to help these students catch up to their grade's curriculum;
6. Students will be more relaxed due to the elimination of high-stakes exams;
7. Students will be more engaged with the curriculum due to the inclusion of non-scholastic dimensions;
8. Student learning outcomes will improve as a result.

We also expect that ABRCs in CCE schools will be able to monitor the program's implementation, and mentor and provide support to teachers.

LEP, on the other hand, is premised on the possibility that, under the status quo, teachers lack the pedagogical tools necessary to teach students according to their ability. Even if teachers are successful in identifying those students who are falling behind grade level, they may be unable to tailor their teaching to meet the needs of low-performers because of both heterogeneous classes and a lack of skills to teach remedial material. Consequently, these students may fall further and further behind as the grade-level curriculum advances. By training teachers to teach to their students' actual levels of skill, LEP gives teachers a methodology—class restructuring according to ability—and a curriculum that emphasizes remedial skills, improving learning outcomes for students who have fallen behind.

Key assumptions of the LEP intervention include:

1. Teachers lack the ability and/or resources to teach their students at their respective levels under the status quo;
2. Teachers will attend trainings for LEP, where they will understand the program's goals and methodology;

3. Teachers and school administrators are willing to put the program in place in their schools;
4. Teachers and headmasters have sufficient time and resources to implement LEP;
5. Teachers will:
    a. Test students at the beginning of the academic year;
    b. Use students' performance on this test to divide them into distinct ability groups (*mahal*s);
    c. Restructure all students in the school by these groups for an hour a day; and
    d. Teach each group using the prescribed Pratham curriculum.
6. The Pratham curriculum is effective in addressing each *mahal*'s skill deficit(s);
7. Student learning outcomes will improve as a result of these *mahal*-based classes.

Once again, we also expect that ABRCs in LEP schools will effectively monitor the implementation of the program, and that they are both willing and able to mentor teachers in putting LEP into place in each treatment school.

The theories of change for both programs, as well as primary and secondary outcomes for each intervention, are detailed in Figure 1 and Tables 2 and 3.

Figure 1: Theory of change

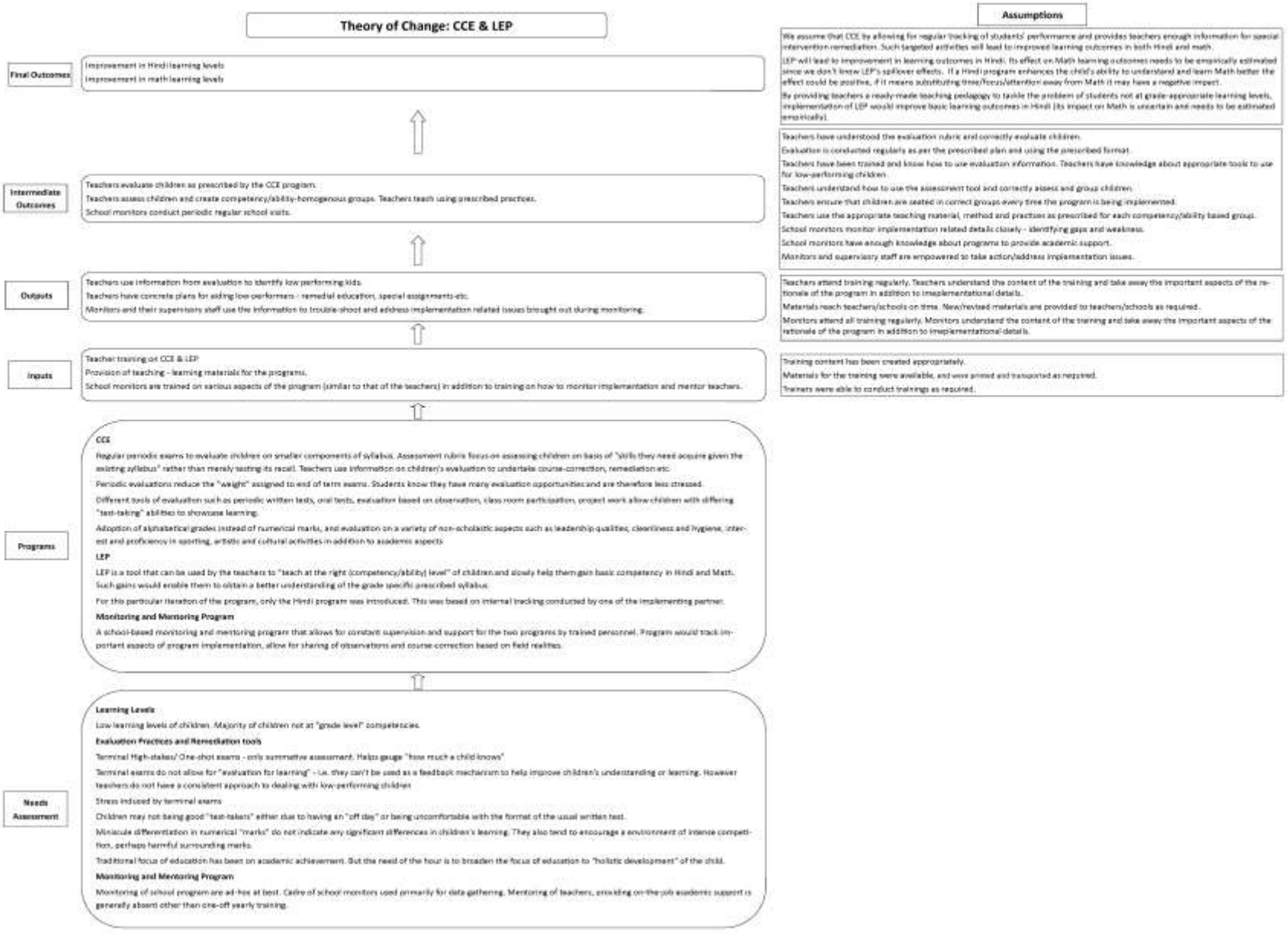


Table 2: CCE theory of change indicators

CCE	Objectives Hierarchy	Indicators	Sources of Verification
Impact (Goal/Overall objective)	Improved student learning outcomes.	Student performance on oral and written tests of basic Hindi and arithmetic.	Endline student testing.
Outcome (Project Objective)	Teachers change teaching practices to address low-performing students.	Teachers' classroom practices.	Process monitoring data (class observation).
Outputs	Teachers gain knowledge of students' abilities through frequent evaluation.	Frequency and methods of student evaluation. Teacher knowledge of student abilities.	Process monitoring data (teacher interviews).
Inputs (Activities)	Teachers and ABRCs trained in CCE methodology.	Number of teachers/ABRCs trained.	Reports from training monitors.

Table 3: LEP theory of change indicators

LEP	Objectives Hierarchy	Indicators	Sources of Verification
Impact (Goal/Overall objective)	Improved student learning outcomes.	Student performance on oral and written tests of basic Hindi and arithmetic.	Endline student testing.
Outcome (Project Objective)	Teachers use Pratham curriculum to teach *mahal*s at the right level.	Teachers' practices in remedial classes.	Process monitoring data (class observation).
Outputs	Teachers administer LEP assessment; restructure classes by *mahal*s for daily remedial classes.	Remedial classes taught/students restructured for class by *mahal*s.	Process monitoring data (class observation).
Inputs (Activities)	Teachers and ABRCs trained in LEP methodology.	Number of teachers/ABRCs trained.	Reports from training monitors.

4. Evaluation design

To estimate the impact of the CCE and LEP programs, we use a randomized-controlled-trial 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 500 rural schools—400 primary schools (grades 1–5) and 100 upper primary schools (grades 6–8) were randomly drawn from a list of all government schools in the four blocks (467 primary and 265 upper primary schools).[1]

Following baseline data testing conducted during the 2011–12 academic year, the 400 primary schools were randomly assigned to one of four treatment groups:

- Group 1 schools (100 schools) received CCE alone
- Group 2 schools (100 schools) received LEP alone
- Group 3 schools (100 schools) received both programs (CCE and LEP) simultaneously
- Group 4 schools (100 schools) received neither program (the control group)

At the same time, the 100 upper primary schools were randomly assigned to one of two treatment groups:

- Group 1 schools (47 schools) received CCE alone
- Group 2 schools (53 schools) received no program (the control group)

LEP was excluded from the upper primary schools since it is intended for grades 3–5 alone, whereas CCE is intended for grades 1 through 8.

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. Informal discussions with teachers conducted at the beginning of our study also raised the issue that primary school teachers may report administratively to the principal of an associated upper primary or high school. It was thus possible that a primary school assigned to the control group could adopt the practices of an associated upper primary school assigned to one of the treatment groups (or vice versa). 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.[2]

For additional details on sample design and random assignment of schools to treatment groups, see Annex A.

[1] Power calculations indicated that a sample of 400 primary schools and 100 upper primary schools would be sufficient to detect a statistically significant improvement in learning outcomes; see Annex B for details.

[2] This randomization at the school campus level is responsible for the uneven assignment of upper primary schools to Group 1 (47 schools) and Group 2 (53 schools).

5. Data

5.1 Data sources

Since, according to GoH, the primary objective of both CCE and LEP was improvement in students' learning outcomes, our primary source of data is students' scores on a series of age-appropriate tests of Hindi and math skills. (Although CCE did include non-scholastic components, GoH considered these to be second-order.) These tests were conducted in two rounds: 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 (400 total), 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 primary school sample of 12,576 students.[1] In upper primary schools (100 total), our sample consisted of all students in 7th grade in each school at baseline, for a total of 3,262 students. For all students in the sample, we collected basic demographic data—including gender, age, and parents' occupations—as well as records of recent school attendance from school registers in each round of testing. Collection of this administrative data was authorized by GoH at the state and district level, and by the headmasters of individual schools in our sample.

At both baseline and endline, all primary school students were administered an individual oral test of basic Hindi and math skills. These tests, developed by the ASER Centre, Pratham's research arm, for use in its Annual Status of Education Report, have been validated in comparison with international testing tools (ASER Centre 2013). The ASER assessment tool tests students on competencies which, according to the standard Indian primary school curriculum, they should be able to successfully demonstrate by the end of grade 2 (for Hindi) and grade 4 (for math).

Table 4: Competencies tested on oral exams

Exam:	**Hindi**	**Math**
Competencies tested:	Cannot identify letters ("Beginner")	Cannot identify single-digit numbers ("Beginner")
	Can identify letters	Can identify single-digit numbers
	Can identify words	Can identify double-digit numbers
	Can read grade 1 level text (short paragraphs)	Can perform subtraction with carry-over
	Can read grade 2 level text (brief story)	Can perform division

The Hindi and math oral tests were each graded on a scale from 0 to 5 at baseline and from 0 to 6 at endline.[2]

[1] In many of the 400 primary schools in our sample, there were fewer than 10 children in certain grades. In cases such as these, all of the children in the grade in question were sampled, yielding a total sample size of less than 10*4*400=16,000 students.

[2] For the endline, the oral Hindi test was modified to include two questions designed to elicit whether students were capable of comprehending passages in addition to being able to read them fluently. These two additional questions account for the extra point at endline; see Annex F for details regarding test scoring.

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 a previous evaluation of the latter's "Read India" program in Bihar and Uttarakhand (Banerjee *et al.* 2011), 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.

Students in upper primary schools were assessed using Hindi and math exams developed by the National Council of Educational Research and Training (NCERT), a national-level education organization providing assistance and advice to state and central education authorities (NCERT 2013). Each exam was scored on a scale from 1 to 60 according to scoring and weighting criteria set by NCERT.

All tests were administered by J-PAL staff during class hours within sample schools. Annex F provides detailed information on scoring criteria for all primary school tests, as well as examples of assessment tools.

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, perception and knowledge of CCE and/or LEP practices (as appropriate), and involvement of ABRCs.

Finally, we incorporated an extensive program of process monitoring into our study design. This monitoring consisted of two 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 CCE and LEP 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 30 minutes to collect data on teaching and evaluation practices in the classroom.

5.2 Sample selection and baseline testing

Following the establishment of partnership between GoH and J-PAL South Asia through the signing of a Memorandum of Understanding in July 2010, GoH decided to engage J-PAL to evaluate their CCE program, which was to be rolled out beginning in the 2011–12 school year. Site selection for the evaluation took place in February and March 2011. An initial round of testing of all students in grade 5 in 200 primary schools and all students in grades 7 and 10 in 100 upper primary schools was undertaken in April and May 2011, at the start of the academic year. These tests for grade 7 constitute the baseline for the upper primary schools in our study.

Upon the completion of this initial round of student testing, several important changes were made to the design of our study. First, due to delays in GoH's planned rollout of CCE, J-PAL and GoH decided to delay the evaluation to the 2012–13 academic year. This led us to drop grade 5 from our design, since in 2012–13 these students would be scattered across a number of upper primary schools, and the intensive tracking needed to prevent large-scale attrition would be prohibitively costly. Finally, to address the

possibility that appropriate pedagogical tools are necessary for CCE to be effective and to address the low learning levels across grades that the baseline testing revealed, J-PAL and GoH decided to add LEP to the study design, and to expand the study sample to include primary school students from all grades.

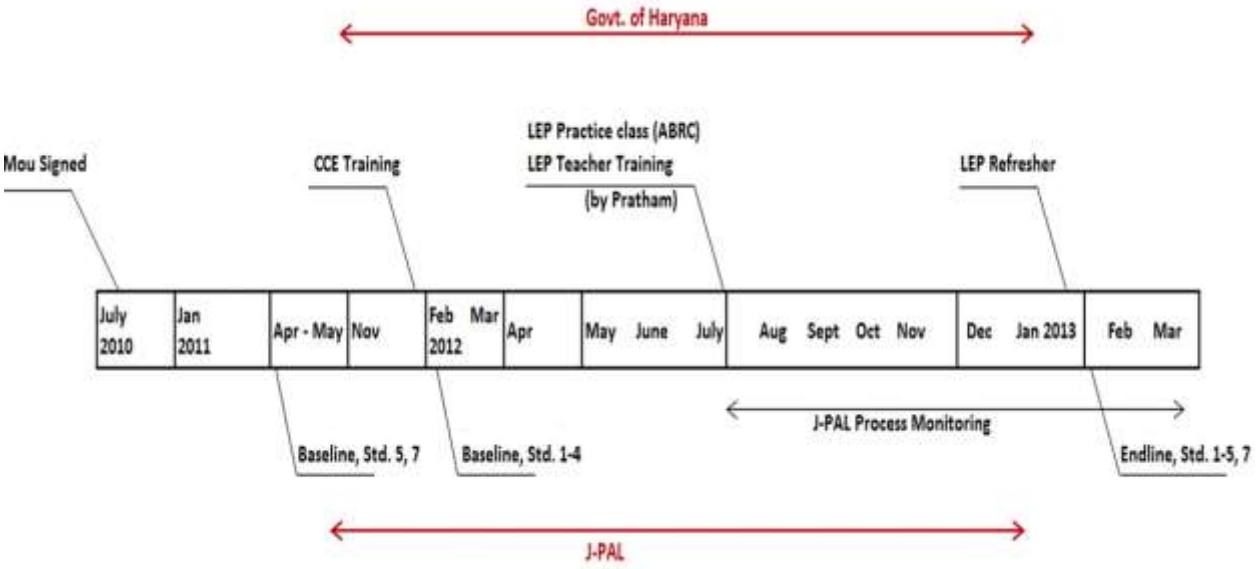
In November 2011, J-PAL South Asia conducted baseline oral testing of up to seven randomly selected students in each of grades 1 to 4 in the 200 primary schools in our original sample, for a total of 4,659 students (if fewer than seven students in a class were enrolled, all students in the class were tested). These 200 primary schools, along with the 100 upper primary schools, were then randomly assigned to one of the four treatment arms of our study. However, due to considerations of statistical power, we decided, following the November baseline, to increase our sample size. To this end, another 200 primary schools were added to the sample, bringing our total to 400. In February and March 2012, up to 10 randomly selected students from each of grades 1–4 in these "new" schools were tested using the oral tests, while written tests were administered to those students who were selected and who were in grades 3 and 4. In the 200 "old" schools, up to three additional students in each grade were given the oral tests to bring the maximum total number of children tested in each grade to 10; of the students tested in either the first or second baselines, all those in grades 3 and 4 were also tested using the written assessments at this time. In all, 7,917 students were added to the sample in this second round of baseline testing, bringing the total primary school sample to 12,576 students. The 200 "new" schools were randomly assigned to one of the four treatment arms following this round of tests.

5.3 Endline testing

Endline student testing was conducted in February and March 2013, at the end of the 2012–13 academic year. Students in all 500 upper primary and primary schools were tested at this time, and schools were visited multiple times to minimize attrition due to student absences. Between primary and upper primary schools, attrition was relatively low: we were able to reach 95.3 per cent of primary school students and 92.0 per cent of students tested at baseline. Statistical tests indicate that attrition was not significantly predicted by students' treatment assignment (see Tables C1 and C3, Panel C). (Table C4 provides additional information on the correlation of student attrition with baseline observable characteristics. Table C5 shows that the composition of attriters is not significantly different between treatment groups at the 10 per cent level.)

6. Program implementation

Figure 2: Timeline of program implementation (not including ABRC training)



6.1 CCE implementation

6.1.1 Teacher training

The first training sessions for CCE took place in November 2011 for teachers in those schools among the 200 schools initially selected for inclusion in the study that had been assigned to receive CCE. However, actual implementation by the teachers did not begin until April 2012, at the same time as the training for teachers in those schools among the 200 "new" schools that had been assigned to CCE. There were a variety of reasons cited by teachers for the delay in implementation in the first set of schools, among them a lack of CCE materials for students (such as report cards and evaluation sheets), uncertainty regarding government guidelines, and unwillingness to introduce a major change in evaluation in the middle of an ongoing academic year.

The GoH partnered with two private-sector education training companies to create the content and deliver the teacher training. A "cascade" model was used for the training: the content creators from these companies trained a group of master trainers who then trained the trainers, who then, in turn, trained the teachers. To ensure maximum participation by teachers, their training was held at the easily accessible block headquarters. The teachers were trained on the CCE program—including how to conduct regular evaluations of students and how to maintain records of student progress—for seven days, and were provided with a CCE manual.

Our process monitoring data show that just over 75 per cent of teachers in schools assigned to implement CCE attended training. After seven days of training, 72 per cent of teacher attendees could successfully identify the main features of CCE, and less than half could successfully list the contents of student report cards (a key component of the program).

6.1.2 Teacher perspectives

Surveys of teachers conducted at the end of CCE training reveal that, on the whole, teachers believed that the program would help students, although about 30 per cent believed that the program would make their jobs more burdensome.

Table 5: Selected teacher perspectives on CCE

	% "Disagree"/"Strongly Disagree"	% "Agree"/"Strongly Agree"
CCE makes teachers' jobs more burdensome.	51.64	28.96
CCE will help the teachers understand the learning level and progress of students better, and hence plan their teaching.	3.57	87.09
CCE will help students who are performing poorly catch up with the others.	3.84	86.30

6.1.3 Program implementation

In our final process monitoring visits to CCE schools, 88.7 per cent of teachers who had received CCE training reported having their CCE manuals; only 42.2 per cent of teachers who had received CCE training were able to show their manual to the surveyor. Likewise, 81.7 per cent of teachers in CCE schools reported using evaluation sheets and 64.7 per cent reported using report cards; only 45.2 per cent and 38.6 per cent of teachers were able to show the surveyor a completed evaluation sheet or a completed report card, respectively, for one of their students.

Although we lack extensive data on the use of specific teaching practices within CCE schools, we do have limited results from classroom observation by J-PAL process monitors. Table 6 shows, for each group of schools, the mean propensity of an observed teacher to use several pedagogical techniques emphasized as part of teacher training for CCE.

Table 6: Use of CCE-encouraged teaching practices

	Primary Schools			Upper Primary Schools		
	Control	*CCE*	*P-value of difference*	*Control*	*CCE*	*P-value of difference*
Teacher uses examples from everyday life	0.490	0.516	0.721	0.426	0.325	0.341
Teacher uses local information to explain concepts	0.150	0.126	0.634	0.234	0.175	0.504
Teacher repeats the concept she/he has just taught based on student answers (without changing practices)	0.020	0.011	0.594	0.064	0.050	0.785
Teacher changes teaching practice (making the explanation simpler) based on student answers	0.420	0.305	0.097	0.383	0.325	0.579

On the whole, teachers in CCE schools do not use the CCE-recommended techniques any more than teachers in control schools. In the one case where we observe a significantly different use of a technique (at the 10 per cent level), CCE teachers in primary schools actually simplify their explanations *less* frequently than control schools. From this limited data, then, it appears that CCE training did not, in itself, lead to changes in teaching practices.

Taken together, these figures indicate that although a majority of teachers in CCE schools were trained in the program, both the absorption of key information by teachers and the actual implementation within classrooms of practices fundamental to CCE were limited. When school headmasters were asked at the end of the 2012–13 year whether they had issues or problems with the CCE program, only 35 per cent said no. Of those who said that they had problems, the most commonly cited issues were feeling overburdened by the additional requirements imposed by CCE, feeling that the program requirements were too time-consuming, and believing that the guidelines for CCE were unclear.

6.2 LEP implementation

6.2.1 Teacher training

As noted above, implementation of LEP in Haryana focused on basic reading skills, and not basic math skills. Teacher training for LEP took place in July 2012. At these sessions, which were conducted by Pratham staff, teachers were introduced to the LEP methodology and were taught how to conduct the simple oral assessments of reading skills; to place students into groups by ability level; and to use the Pratham remedial curriculum to teach students in each group. Teachers were also provided with materials

for LEP implementation, including manuals and classroom materials (such as flashcards, story books, and worksheets) to use in the LEP remedial classes. A critical component of Pratham's training was a field training session in which teachers were taken to nearby schools to practice both running assessments and teaching remedial classes.

In November 2012, Pratham staff conducted an internal midline assessment, after which they reported to J-PAL that they believed that students were not progressing sufficiently in the remedial curriculum. As a result, Pratham conducted a second round of "refresher" trainings in January 2013 to reinforce teachers' knowledge of LEP practices with respect to reading. Importantly, Pratham also decided, following their midline assessment, to conduct a more concerted and intensive push for implementation of the Hindi curriculum over the math curriculum. Thus, implementation of LEP in schools in our sample was focused solely on grouping students by Hindi ability and teaching remedial Hindi classes, rather than on teaching both remedial Hindi *and* remedial math classes according to ability level.

6.2.2 Teacher perspectives

Surveys of teachers conducted at the end of LEP training reveal that, on the whole, teachers believed strongly that the program would help identify skill deficits among students. Teachers were fairly evenly split on the difficulty of implementing the regrouping of students, although less than 20 per cent believed that the program would make their jobs more burdensome.

Table 7: Selected teacher perspectives on LEP

	% "Disagree"/"Strongly Disagree"	% "Agree"/"Strongly Agree"
LEP tools will make it easier to identify where students are performing poorly	2.91	93.82
Regrouping students by competence every day will be difficult	44.80	40.89
The LEP program makes teachers' jobs more burdensome	69.44	18.89
Devoting time and resources to LEP classes will detract from the learning of higher-performing students	45.57	41.49
Relative to current practices, LEP will be more effective at improving the learning levels of low-performing students	3.27	88.93

6.2.3 Program implementation

On the whole, implementation of LEP was more successful than implementation of CCE. Data from process monitoring show that, in primary schools assigned to LEP, 93.9 per cent of teachers attended a Pratham training. In our final process monitoring visits to LEP schools, 98.1 per cent of teachers who had received LEP training reported having their LEP manuals, and 64.6 per cent of teachers who had attended LEP training were

able to show their manual to the surveyor. Eighty-five per cent of LEP schools had completed assessment sheets for their students, and headmasters in 99.4 per cent of LEP schools reported that they conducted remedial classes every day. Data from direct observation of these classes indicated a high degree of compliance with the Pratham curriculum by teachers: students were taught according to mahals—the pivotal component of LEP—in 94.0 per cent of remedial classes observed, and approximately 70 per cent of the interactive games that teachers played with students in remedial classes were appropriate to the students' level according to the LEP curriculum.

6.3 ABRC implementation

A total of 42 ABRCs received training in CCE in April 2012 and in LEP in May 2012. As with the teacher training sessions, the CCE training was conducted by a private-sector education training firm, and the LEP training was conducted by Pratham (7 days for CCE, 18 days for LEP). In August 2012, ABRCs were trained by J-PAL South Asia on mentoring and monitoring for both programs, as well as on general monitoring and mentoring practices (a total of four days). This training also contained a unit on randomized evaluation in order to familiarize the ABRCs with the study design and to explain to them the importance of non-contamination across treatment arms. ABRCs were instructed to visit schools on a regular basis, spending at least two hours in the school checking records, observing a class in session, and providing on-site mentoring and guidance, and filling out a monitoring data tool with questions on resource availability, teaching practices, program implementation, and general documentation.

In general, implementation of the ABRC mentoring and monitoring was thorough. The frequency of ABRC visits to schools varied from month to month, but at our final process monitoring visit to each school, 80 per cent of schools reported a visit from an ABRC in the previous 30 days. Of those who reported a visit, 76.5 per cent said that the ABRC spent over an hour in the school, and 93.1 per cent said that the ABRCs observed a class in progress on at least one visit. Teachers in treatment schools made use of the ABRCs as resources for program implementation, especially for CCE: 53.9 per cent of teachers in LEP schools reported asking ABRCs questions about LEP implementation, with 97.6 per cent of those asking questions reporting that the ABRC's answer was helpful, and 71.2 per cent of teachers in CCE schools reported asking ABRCs questions about CCE implementation, with 94.8 per cent of those who asked questions reporting that the ABRC's answer was helpful.

6.4 Fidelity to protocols

Perhaps the greatest challenge to our evaluation design was the possibility of contamination in the non-CCE treatment arms. As a large-scale national program, CCE has been touted by state and national educational institutions, and has frequently appeared in the Indian news media. The possibility also existed of contamination of LEP into non-LEP schools. Primarily because teachers in our study area were previously aware of CCE, multiple steps were taken to ensure that schools implemented only those programs to which they had been assigned. At the start of the 2012–13 year, GoH sent letters to schools informing them that new programs were to be rolled out in a staggered manner in Haryana, and instructing them to wait until their teachers were trained in any new program before implementing it. Although this letter did make schools aware of the existence of other programs, the research team felt it would be preferable to send the

letter and make it clear to schools that they were not to change their teaching practices in the absence of a mandate to do so from GoH. Careful monitoring of implementation by ABRCs and J-PAL process monitors was also used to ensure fidelity to the random assignment, and ABRCs were trained on how to avoid inadvertent contamination through, for example, suggesting that teachers in one treatment arm utilize aspects from another treatment arm's intervention in their teaching.

Our data suggest very low levels of contamination across treatment arms. Just 0.82 per cent of teachers in non-CCE schools were trained in CCE, and 0.38 per cent of teachers in non-LEP schools received training in LEP. Data on contamination by ABRCs is similar: only one non-CCE school reported that an ABRC had discussed CCE during a visit, and no non-LEP schools reported that ABRCs discussed LEP during a visit.

7. Results

7.1 Baseline testing

Baseline test results are summarized in Tables C1, C2, and C3. Results from baseline oral tests are shown in Figures 3 and 4. Consistent with recent statewide surveys (ASER Centre 2011b, 2013), we find that competencies in both Hindi and math in our sample were generally poor. Over 25 per cent of primary school students in our sample were able to identify isolated letters, and almost 84 per cent of students were unable to read a simple story of 8–10 lines rated for grade 2. More than 55 per cent of primary school students tested were unable to recognize two-digit numbers. Following the patterns from the statewide ASER survey (ASER Centre 2013), students in Mahendragarh tended to perform better on average than students in Kurukshetra; for example, 22.5 per cent of students in Mahendragarh were able to read a grade 2 text, compared with just 12.2 per cent in Kurukshetra.

Figure 3: Hindi language competencies in primary schools, baseline

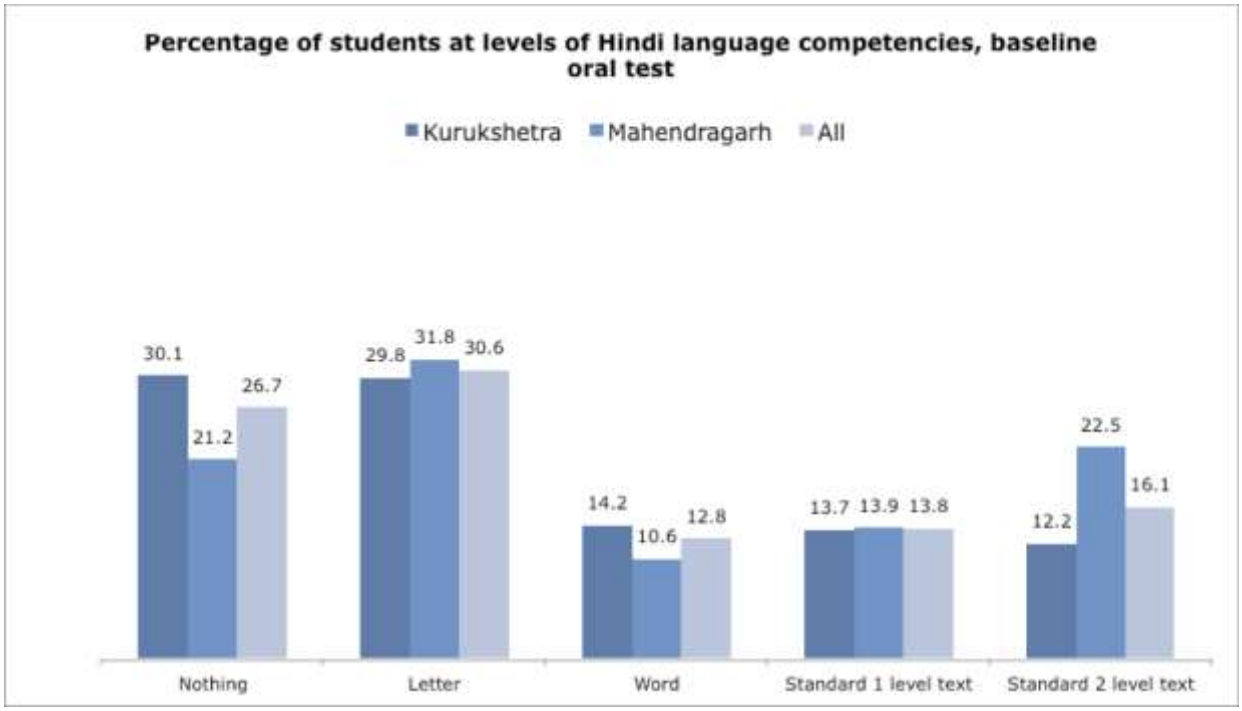
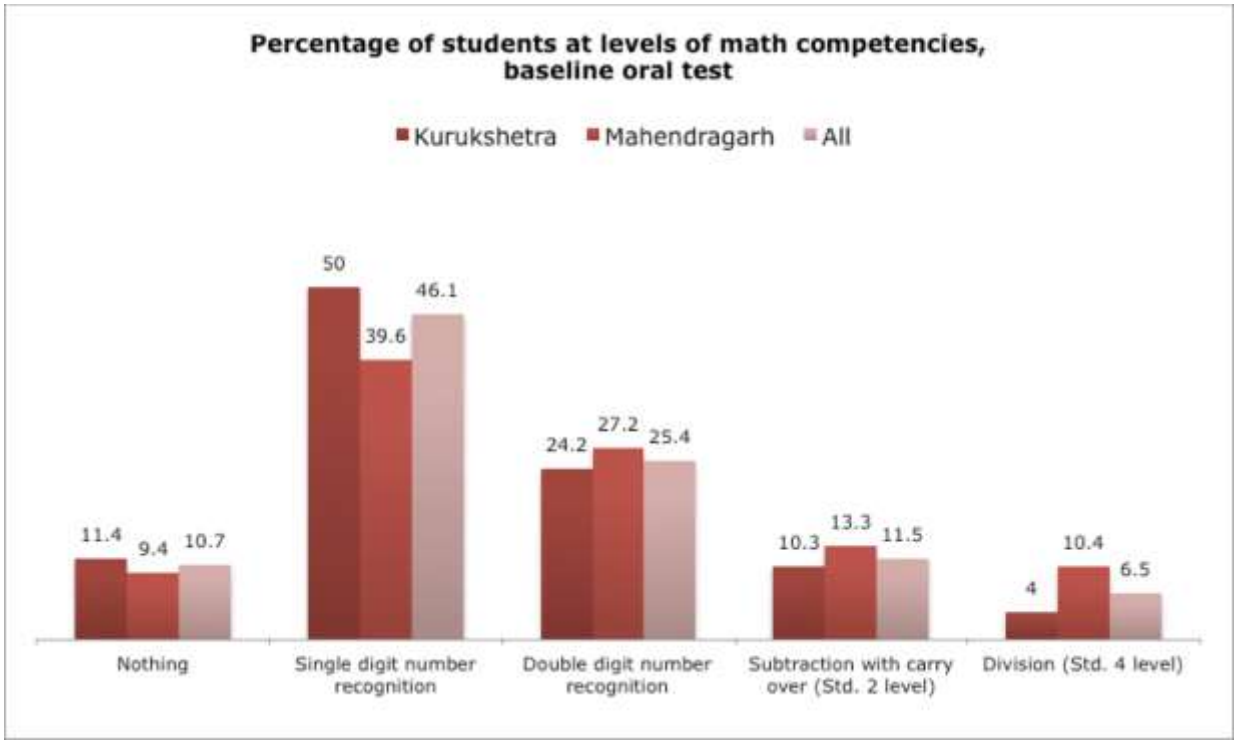


Figure 4: Math competencies in primary schools, baseline



Figures 5 and 6 also show baseline oral Hindi and math results, respectively, broken down by treatment group.

Figure 5: Hindi language competencies in primary schools, baseline

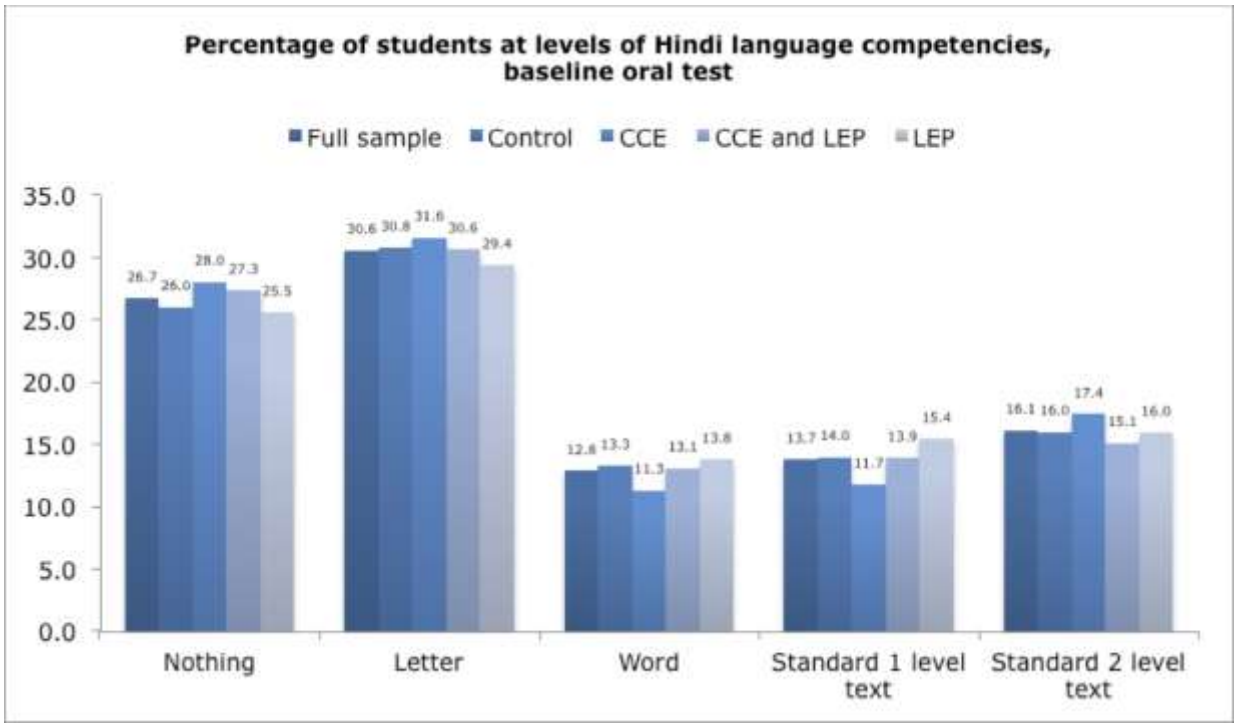
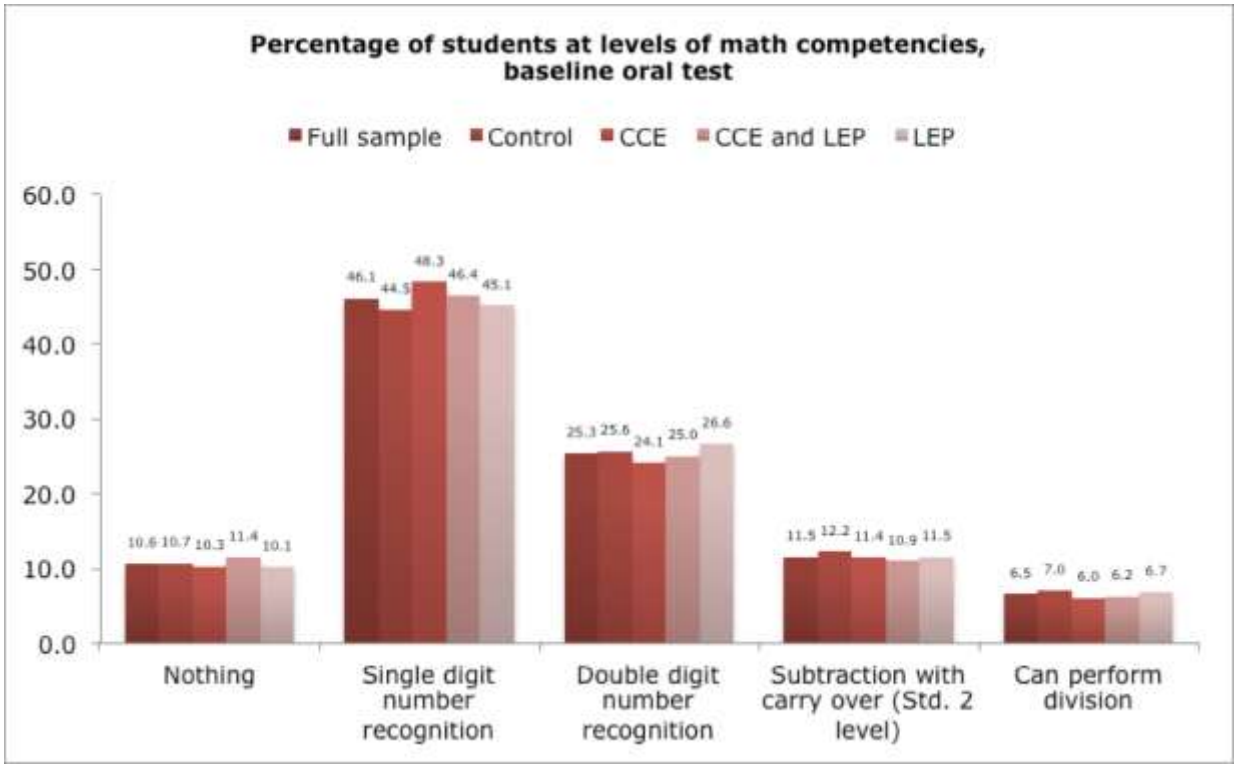


Figure 6: Math competencies in primary schools, baseline



Both the written tests administered to students in grades 3 and 4 at baseline and the NCERT-developed tests administered to grade 7 students showed similar patterns of poor performance. The mean written Hindi test score for students in grades 3 and 4 was 4.04 out of 12.5, and 76 per cent of students in our sample received half or fewer of the total possible points on the test. For math, the mean score for students in grades 3 and

4 was 5.01 out of 12.5, and 64 per cent of students received half or fewer of the total possible points. For upper primary students, the mean Hindi test score was 19.0 out of 60 possible points and the mean math test score was 14.1 out of 60. Almost 90 per cent of 7th graders tested scored a 50 per cent or worse on the Hindi test; this number jumps to 99.5 per cent of students for the math test.

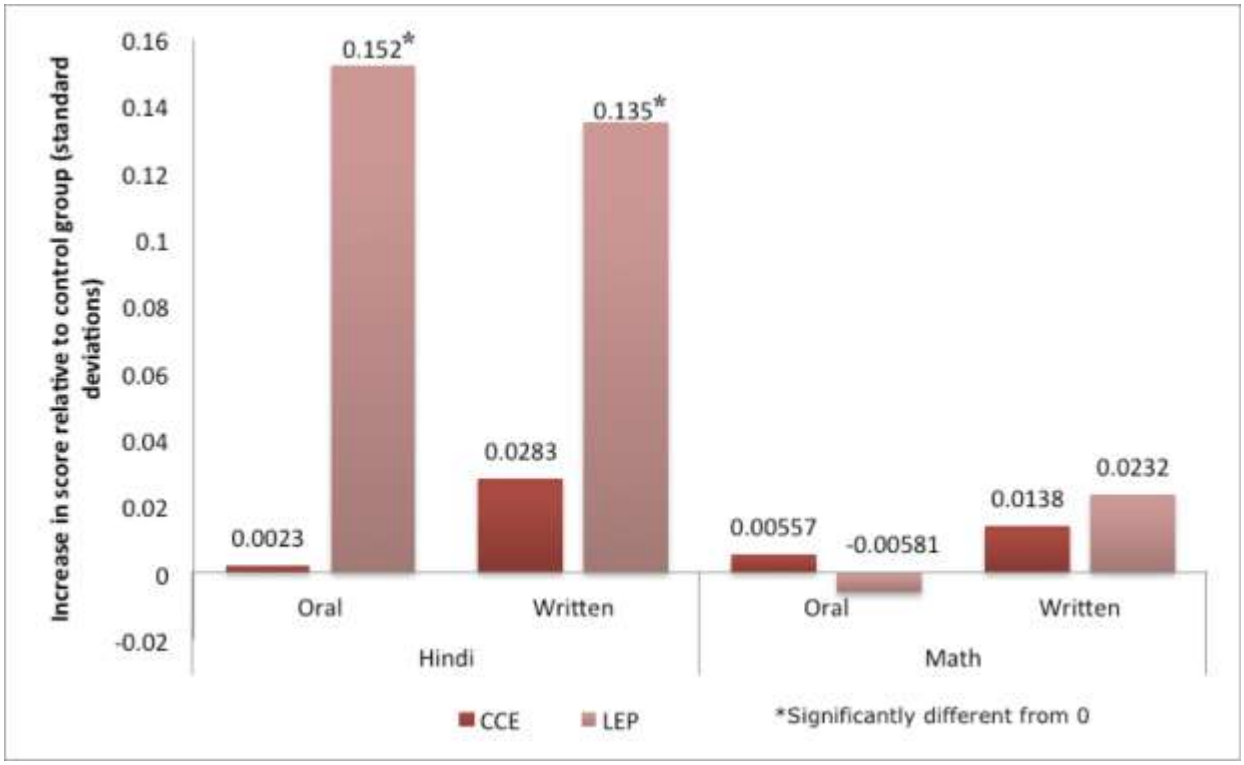
7.2 Impact results

We present impact results for primary schools first, followed by results for upper primary schools.

7.2.1 Primary schools: average treatment effects

To examine the effects of the LEP and CCE programs (both individually and in combination) on learning outcomes, we regress students' test scores at endline on indicators for each program, controlling for baseline test scores, age, grade in school, and gender (Table E1). In these regressions, test scores in each round are normalized using the mean and standard deviation of the control group in that round.
The CCE program had no significant effect on either oral or written test scores for either Hindi or math. Being in an LEP school, in contrast, had a large and statistically significant effect on students' Hindi learning outcomes. On the oral Hindi test, students in LEP schools scored 0.15 standard deviations higher on average relative to the control group, corresponding to a boost of 0.3 points out of 6. On the written Hindi test for grades 3–5, students in LEP schools scored 0.135 standard deviations higher relative to the control group, corresponding to an increase of 0.4 points out of 12.5. LEP had no significant effect, however, on either oral or written scores in math. The additional effect of combining CCE and LEP was not significantly different from zero for any test, indicating that LEP did not, as hypothesized, improve the effectiveness of CCE (or vice versa).

Figure 7: Average impact results, primary schools



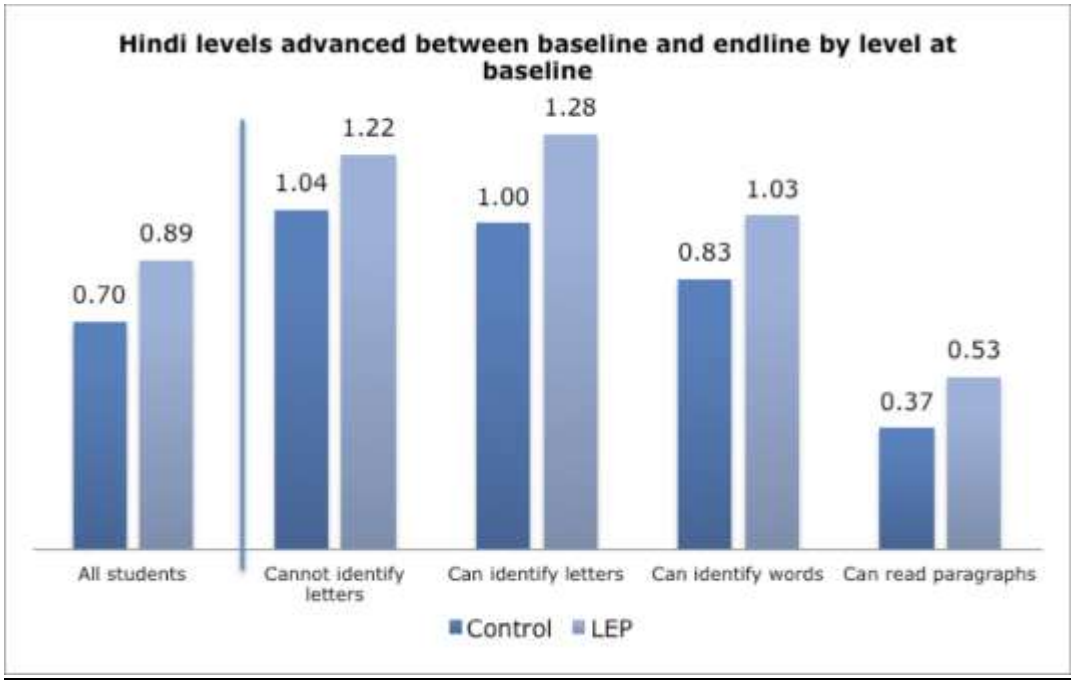
i Primary schools: heterogeneity in treatment effects of LEP

1. Effects by ability level at baseline

To examine LEP's differential effects on different groups of students, we look first at the impact of LEP on students' progress through the competencies identified on the oral Hindi test by students' ability level at baseline (i.e. Beginner, Letter, Word, Paragraph, Story). Figure 8 shows the average number of categories advanced by students in control and LEP schools; these results are from a regression of the levels moved by a student on treatment. Note that a one-level jump does not represent an "equal" jump in ability at each level; in other words, moving from "beginner" to "letter" cannot be said to be quantitatively equivalent to moving from "word" to "paragraph." Thus, these results should be taken to be suggestive.

As Figure 8 shows, all categories of students advanced a greater number of levels between baseline and endline in LEP schools relative to their peers in control schools.

Figure 8: LEP impact on Hindi competencies, by baseline Hindi level

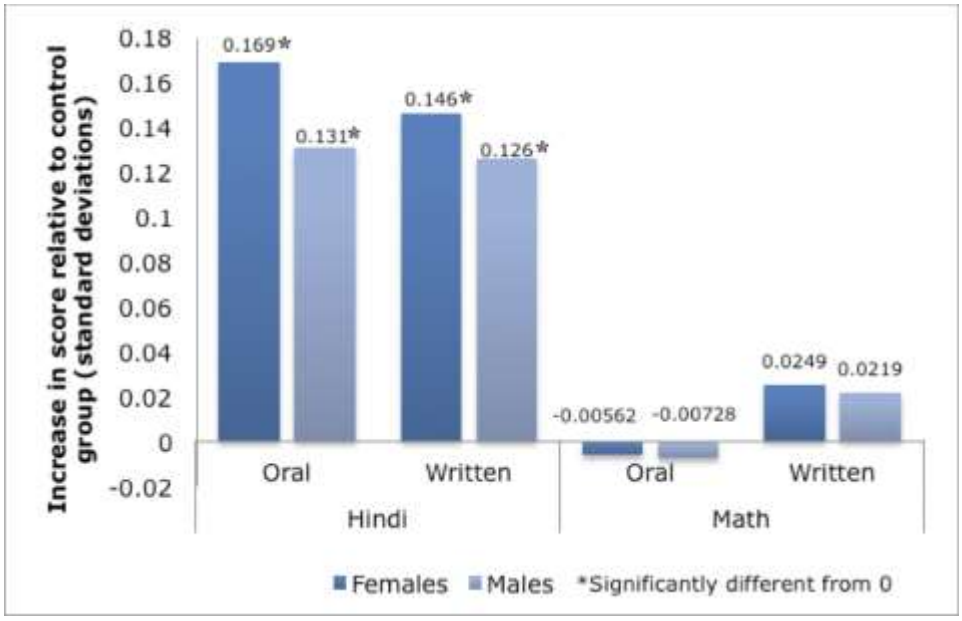


(Note: LEP-Control differences at every level are significantly different from zero.)

2. Effects by gender

Table E2 shows the impact of treatment on students' test scores at endline, separately for boys and girls, controlling for baseline test scores, age, and grade. Although LEP had a large and significant impact for both genders, it had a larger effect for girls than boys on both oral Hindi test scores (0.169 versus 0.131 standard deviations relative to control, significantly different at the 10 per cent level) and written Hindi test scores (0.146 versus 0.126 standard deviations relative to control, not significantly different).

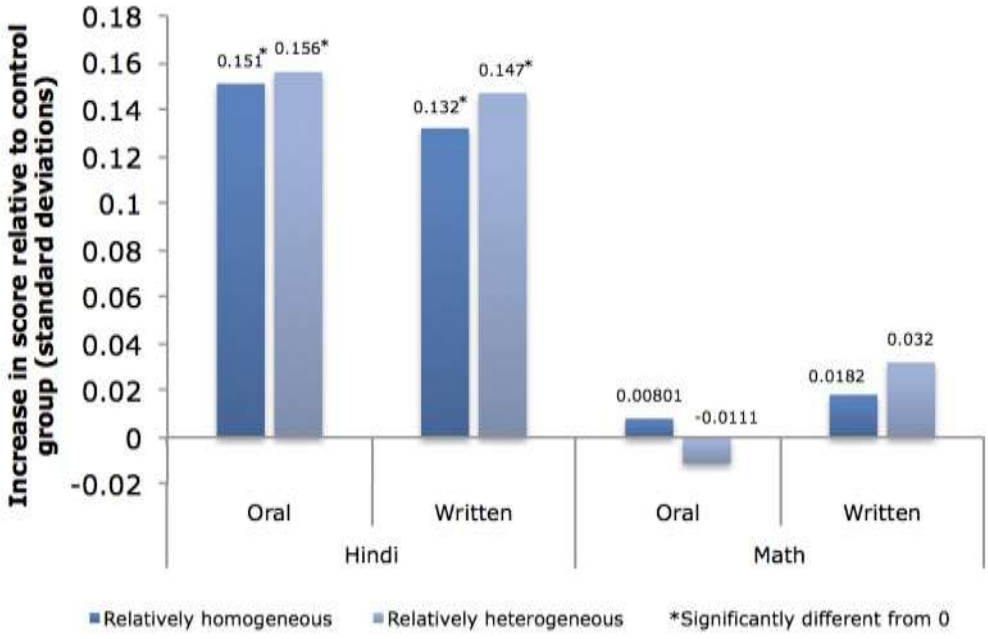
Figure 9: LEP impact results by gender, primary schools



7.2.2 Effects by in-class heterogeneity

A key component of LEP's theory of change is that the program allows students to be better taught at their own level, something which may be difficult to do in classrooms with large variation in student ability. In Table E3, we therefore examine the impact of LEP on student test scores at endline in relatively homogeneous classes versus relatively heterogeneous classes. Heterogeneity was measured by the variance in baseline scores for each test, with high/low heterogeneity classes defined as those with variance above/below the median for all classes in our sample. As Figure 10 shows, LEP had a large and significant positive impact for both groups of students.

Figure 10: LEP impact by baseline in-class heterogeneity in learning levels, primary schools



7.2.3 Upper primary schools

To examine the effects of CCE on learning outcomes, we regress upper school students' test scores at endline on an indicator for assignment to the treatment group, controlling for baseline test scores, age, and gender (Table E4). As with the regressions for primary school students, test scores in each round are normalized using the control group mean and standard deviation. In our sample of 7th graders, students in CCE schools did not perform significantly better than students in control schools on either the NCERT Hindi or NCERT math test.

8. Policy recommendations

The LEP program's large effect on students' basic Hindi skills indicates that programs emphasizing teaching at the right level can play an important role in improving poor learning outcomes in developing-country contexts. Our results for the impact of LEP are similar to those of previous studies evaluating the impact of Pratham's teaching at the right level curriculum (such as Banerjee *et al.* 2011). What is unique in the case of our study is that the TaRL pedagogy was implemented by government teachers and not Pratham volunteers or staff. Our findings thus provide strong evidence for the positive impact of teaching at the right level even when integrated into a formal school system, and when operated on a large scale.

Students in LEP schools, however, did not perform any better than students in control schools on oral or written tests of basic math skills. Given that mathematics was not explicitly included in LEP in the 2012–13 school year, this indicates that the program had no spillovers (positive or negative) from reading to math. These results suggest that it would be important to explicitly cover both reading and math components when implementing LEP in the future, to ensure that any potential gains in student achievement are not restricted to reading alone.

It is also important to note that a critical component of LEP as it was implemented in our study was the monitoring and mentoring provided by ABRCs. Given the high use of ABRCs as sources of information about LEP practices by teachers in LEP schools, it seems likely that future versions of LEP which fail to include careful monitoring should not necessarily expect to find impacts similar to those that we find.

We find that the CCE had no impact on student learning outcomes in Hindi or math for either primary or upper primary school students. Coupled with our evidence of CCE's relatively mediocre implementation, these results suggest that the CCE scheme in its current form needs a thorough review in design and appropriateness, as well as in teacher training and implementation. Regular evaluation of pupils is essential to teaching, but the complexity of CCE's evaluation tools and the lack of a clear connection between such evaluations and specific changes in teaching practices may have limited the usefulness of CCE as it was implemented in our setting. It is conceivable that if CCE focuses on basic and foundational skills rather than on standard-level competencies, it may be more effective.

Appendix A: Sample and study design

The study area consisted of two districts in Haryana, Mahendragarh, and Kurukshetra, which were selected by GoH and J-PAL because of their different educational and economic profiles. Within each district, two blocks were chosen at random for inclusion in our study: Ateli and Narnaul in Mahendragarh, and Pehowa and Thanesar in Kurukshetra. Schools were randomly selected from a list of all primary and upper primary schools in these four blocks.

Table A1: Schools selected for sample, by block and school type

District		**Mahendragarh**		**Kurukshetra**		
Block		Ateli	Narnaul	Pehowa	Thanesar	*Total*
School type	Primary	89	76	101	134	400
	Upper Primary	19	18	24	39	100
Total		108	94	125	173	500

In the 400 primary schools, the sample consisted of a total of 12,576 students in grades 1–4 in the 2011–12 academic year. In each school, up to 10 students in each of these four grades were randomly selected from the school's enrollment register for baseline testing; if fewer than 10 students in a class were enrolled, all students in the class were tested. Baseline testing took place in two rounds. In the first (November 2011), up to seven students in each grade in each of 200 of these primary schools were tested; in the second (February/March 2012), up to 10 students in each grade in the 200 remaining schools were tested, and up to three additional students were tested in each grade in the original set of 200 primary schools.

In the 100 upper primary schools, the sample consisted of 3,262 students, all of whom were in 7th grade during the 2011–12 academic year. Unlike in the primary school sample, we did not select a random subset of these students for baseline testing. Instead, all 7th grade students in each school were administered tests at baseline.

The distribution of students included in the sample across blocks and school types was as follows:

Table A2: Students selected for sample, by block and school type

District		**Mahendragarh**		**Kurukshetra**		
Block		Ateli	Narnaul	Pehowa	Thanesar	*Total*
School type	*Primary*	2,493	2,302	3,161	4,620	12,576
	Upper Primary	494	558	915	1,295	3,262
Total		2,987	2,860	4,076	5,915	15,838

Random assignment to treatment groups occurred following baseline testing. To prevent spillovers between schools of different levels sharing the same school campus, randomization took place at the school campus level, stratifying for block and mean baseline test scores. Each school campus was assigned to one of four treatment groups: (a) CCE alone, (b) LEP alone, (c) CCE and LEP together, and (d) neither program. Since LEP was implemented only in primary schools, upper primary schools in school campuses

assigned to (a) and (c) constituted the CCE treatment group for the upper primary sample, while those in campuses assigned to (b) and (d) constituted the control group.

The distribution of schools and students in each treatment arm was as follows:

Table A3: Random assignment of schools and students, by sample and treatment group

		CCE	CCE and LEP	LEP	Control	*Total*
Primary sample	*Schools:*	100	100	100	100	400
	Students:	3,184	3,200	3,175	3,017	12,576
		CCE		**Control**		
Upper primary sample	*Schools:*	47		53		100
	Students:	1,527		1,835		3,262

Spillovers and John Henry/Hawthorne Effects

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. Informal discussions with teachers conducted at the beginning of our study also raised the issue that primary school teachers may report administratively to the principal of an associated upper primary or high school. It was thus possible that a primary school assigned to the control group could adopt the practices of an associated upper primary school assigned to one of the treatment groups (or vice versa).

Due to fears of spillovers between schools sharing a campus, randomization was conducted on the level of the "super-school," or 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. Thus, all schools sharing a school campus were assigned to the same treatment, mitigating possible spillovers. Since LEP was implemented only in primary schools, upper primary schools in school campuses assigned to "CCE" and "CCE and LEP" constituted the CCE treatment group for the upper primary sample, while those in campuses assigned to "Control" and "LEP" constituted the control group.

Another key concern was the potential for spillovers between treatment arms as a result of teachers learning of programs and choosing to implement them independently. For example, as a large-scale national program, CCE has been touted by state and national educational institutions and has frequently appeared in the Indian news media. Primarily because teachers in our study area were previously aware of CCE, multiple steps were taken to ensure that schools implemented only those programs to which they had been assigned. At the start of the 2012–13 year, GoH sent letters to schools informing them that new programs were to be rolled out in a staggered manner in Haryana, and instructing them to wait until their teachers were trained in any new program before implementing it. Although this letter did make schools aware of the existence of other programs, the research team felt it would be preferable to send the letter and make it clear to schools that they were not to change their teaching practices in the absence of a mandate to do so from GoH.

Careful monitoring of implementation by ABRCs and J-PAL process monitors was also used to ensure fidelity to the random assignment, and ABRCs were trained on how to avoid inadvertent contamination through, for example, suggesting that teachers in one treatment arm utilize an aspect from another treatment arm in their teaching.

More generally, the letter from the Government was intended to dissuade control schools from changing their teaching practices solely as a result of inclusion in the study (John Henry effects). Hawthorne effects were unlikely to be a significant concern because monitoring was a central component of the program itself. (Process monitoring evaluators did visit schools, but these formed a small proportion of the total visits to each school by ABRC monitors, Pratham monitors, and J-PAL monitors alike.) Teachers, however, may have changed their classroom behaviors on the days of the process monitoring visits. While we were unable to check explicitly whether or not this was the case, we verified our classroom observations with schools records whenever possible. Further, the visits by J-PAL process monitors were unannounced, and data such as teacher absence and teacher activity at the beginning of the visits are not affected by this issue.

Appendix B: Power calculations

We base our power calculations on the written test scores, with standard values of 0.80 for power and a significance level of 0.95. For the other parameters of the calculation, we use data from Banerjee *et al.* (2010) on standardized test scores for 3rd and 4th grade children in government schools in Vadodara and Mumbai. Using these data, we estimate an intra-cluster correlation (correlation between students within a school) of 0.17. For the correlation between baseline and endline test scores, we estimate a value of 0.57.

Under these assumptions and using a cluster size of 30 students, we will be able to detect differences between treatment groups in the main sample (the 400 primary schools) of 0.13 standard deviations. Pooling treatment groups that contain a single intervention (e.g. "CCE only" and "CCE and LEP" to test the effects of any CCE intervention) and comparing to control reduces the minimum detectible effect to 0.09 standard deviations. To detect the effects of CCE within the grade 7 sample of 100 schools (CCE vs. control), we require an effect size of 0.18 standard deviations. By comparison, Banerjee *et al.* (2010) find effect sizes ranging from 0.14 to 0.28 standard deviations for the Balsakhi program, which was considered a successful intervention.

Appendix C: Descriptive statistics and balance check of randomization

Table C1: Descriptive statistics and balance check of randomization, primary school sample

	Full sample	Control	CCE	CCE and LEP	LEP	P-value of F-test of joint significance
A. Demographic characteristics						
Female (%)	50.83	50.61	51.07	51.47	50.17	0.875
	(50.00)	(50.00)	(50.00)	(49.99)	(50.01)	
Age (years)	9.058	9.042	9.034	9.058	9.098	0.764
	(1.598)	(1.568)	(1.621)	(1.584)	(1.618)	
Grade in 2011–12 school year	2.554	2.566	2.541	2.549	2.558	0.516
	(1.117)	(1.114)	(1.124)	(1.114)	(1.116)	
B. Baseline test scores						
Baseline oral Hindi test score (out of 6)	2.553	2.571	2.513	2.503	2.626	0.600
	(2.020)	(2.009)	(2.057)	(2.004)	(2.008)	
Baseline oral math test score (out of 6)	2.029	2.063	2.011	2.011	2.033	0.863
	(1.431)	(1.449)	(1.433)	(1.425)	(1.416)	
Baseline written Hindi test score (out of 12.5)	4.041	4.118	4.042	4.024	3.983	0.916
	(2.629)	(2.619)	(2.684)	(2.628)	(2.583)	
Baseline written math test score (out of 12.5)	5.006	5.127	5.145	4.897	4.859	0.561
	(3.373)	(3.406)	(3.440)	(3.348)	(3.292)	
C. Attrition						
Not reached at endline (%)	4.71	5.27	4.24	5.28	4.06	0.317
	(21.18)	(22.35)	(20.15)	(22.37)	(19.75)	
D. Endline test scores						
Endline oral Hindi test score (out of 6)	3.137	3.032	2.931	3.264	3.316	
	(1.990)	(1.952)	(1.997)	(2.009)	(1.976)	
Endline oral math test score (out of 6)	2.945	2.999	2.927	2.931	2.926	
	(1.622)	(1.631)	(1.648)	(1.626)	(1.583)	
Endline written Hindi test score (out of 12.5)	5.066	4.909	4.887	5.251	5.207	
	(2.612)	(2.628)	(2.676)	(2.573)	(2.551)	
Endline written math test score (out of 12.5)	6.262	6.308	6.226	6.277	6.239	
	(3.491)	(3.507)	(3.597)	(3.428)	(3.435)	

Notes: Standard deviations reported in parentheses. The extreme right column reports the P-value of an F-test of joint significance of treatment assignment as a predictor of the variable for that row using an OLS regression.

Table C2: Hindi and math ability levels by round of testing, primary schools (per cent of students)

	Full sample	Control	CCE	CCE and LEP	LEP
A. Baseline oral tests					
Hindi					
Cannot identify letters	26.73	26.01	28.01	27.33	25.54
Can identify letters	30.58	30.78	31.56	30.64	29.35
Can identify words	12.84	13.26	11.29	13.07	13.75
Can read paragraph	13.74	13.96	11.73	13.89	15.41
Can read story	16.11	15.99	17.41	15.08	15.95
		P-value of chi-squared test of independence: 0.000			
Math					
Cannot identify numbers	10.61	10.66	10.25	11.41	10.10
Can identify single-digit numbers	46.08	44.50	48.25	46.41	45.09
Can identify double-digit numbers	25.31	25.63	24.06	25.02	26.56
Can perform subtraction	11.50	12.20	11.40	10.93	11.50
Can perform division	6.50	7.02	6.03	6.24	6.74
		P-value of chi-squared test of independence: 0.122			
B. Endline oral tests					
Hindi					
Cannot identify letters	7.87	8.02	10.05	7.31	6.12
Can identify letters	21.54	22.74	24.21	20.66	18.62
Can identify words	20.91	21.65	20.99	19.01	22.01
Can read paragraph	20.39	20.95	18.36	21.39	20.89
Can read story	29.29	26.63	26.38	31.64	32.37
Math					
Cannot identify numbers	2.05	2.00	2.40	2.12	1.68
Can identify single-digit numbers	29.60	29.50	29.96	30.60	28.35
Can identify double-digit numbers	31.54	29.67	32.23	30.96	33.18
Can perform subtraction	19.44	20.03	17.54	19.56	20.68
Can perform division	17.36	18.80	17.87	16.75	16.11

Table C3: Descriptive statistics and balance check of randomization, upper primary school sample

	Full sample	Control	CCE	P-value of F-test of joint significance
A. Demographic characteristics				
Female (%)	50.97	49.56	52.77	0.632
	(50.00)	(50.01)	(49.94)	
Age (years)	13.888	13.943	13.818	0.084
	(1.181)	(1.182)	(1.176)	
B. Baseline test scores				
Baseline NCERT Hindi score (out of 60)	19.010	19.115	18.873	0.739
	(8.372)	(8.525)	(8.170)	
Baseline NCERT math score (out of 60)	14.121	14.079	14.176	0.787
	(4.481)	(4.609)	(4.311)	
C. Attrition				
Not reached at endline (%)	8.00	7.91	8.13	0.898
	(27.14)	(26.99)	(27.34)	
D. Endline test scores				
Endline NCERT Hindi score (out of 60)	22.324	22.208	22.474	
	(9.727)	(9.794)	(9.642)	
Endline NCERT math score (out of 60)	14.988	15.036	14.926	
	(4.491)	(4.349)	(4.667)	

Notes: Standard deviations reported in parentheses. The extreme right column reports the P-value of an F-test of joint significance of treatment assignment as a predictor of the variable for that row using an OLS regression.

Table C4: Attrition and observable characteristics at baseline

A. Primary schools	Female	Age (years)	Grade in 2011–12 school year	Baseline oral Hindi test score (out of 6)	Baseline oral math test score (out of 6)	Baseline written Hindi test score (out of 12.5)	Baseline written math test score (out of 12.5)
	-0.00595	0.00128	-0.00919***	-0.00157	-0.00207	0.00224**	0.00139*
	(0.00397)	(0.00134)	(0.00181)	(0.00105)	(0.00148)	(0.00108)	(0.000826)
Observations	12,576	12,555	12,576	12,472	12,393	6,208	6,204
R-squared	0.0181	0.0182	0.0203	0.0182	0.0184	0.0227	0.0224

B. Upper primary Schools	Female	Age (years)	Baseline NCERT Hindi score (out of 60)	Baseline NCERT math score (out of 60)
	-0.0326**	0.0224***	-0.000643	-0.000764
	(0.0139)	(0.00547)	(0.000574)	(0.00117)
Observations	3,261	3,255	2,610	2,602
R-squared	0.0303	0.0361	0.0356	0.0327

Notes: In each panel, each column presents the coefficient from a regression of "not found between baseline and endline" on the baseline characteristic shown. Regressions include fixed effects for strata used in randomization (coefficients not shown). Robust standard errors in parentheses (clustered at school campus level).
* Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level.

Table C5: Balance check of composition of attriters

	All attriters	Control	CCE	CCE and LEP	LEP	P-value of F-test of joint significance
A. Primary Schools (n=592)						
Female (%)	48.31	49.69	48.89	50.89	42.64	0.563
	(50.01)	(50.16)	(50.17)	(50.14)	(49.65)	
Age (years)	9.132	9.072	9.232	9.160	9.065	0.874
	(1.747)	(1.655)	(1.853)	(1.727)	(1.784)	
Grade in 2011−12 school year	2.307	2.289	2.370	2.260	2.326	0.846
	(1.096)	(1.110)	(1.098)	(1.087)	(1.098)	
Baseline oral Hindi test score (out of 6)	2.490	2.274	2.608	2.317	2.856	0.117
	(2.019)	(1.977)	(2.059)	(1.970)	(2.055)	
Baseline oral math test score (out of 6)	1.997	1.890	2.050	1.927	2.165	0.500
	(1.479)	(1.462)	(1.542)	(1.451)	(1.471)	
Baseline written Hindi test score (out of 12.5)	4.649	4.690	4.530	4.784	4.567	0.971
	(2.656)	(2.542)	(2.797)	(2.774)	(2.538)	
Baseline written math test score (out of 12.5)	5.668	5.812	5.338	5.867	5.625	0.887
	(3.380)	(3.320)	(3.753)	(3.242)	(3.232)	
B. Upper Primary Schools (n=261)						
Female (%)	44.06	38.62	50.86			0.395
	(49.74)	(48.86)	(50.21)			
Age (years)	14.202	14.231	14.165			0.761
	(1.502)	(1.480)	(1.535)			
Baseline NCERT Hindi score (out of 60)	18.209	18.469	17.954			0.757
	(8.455)	(9.485)	(7.366)			
Baseline NCERT math score (out of 60)	13.984	14.226	13.754			0.616
	(4.634)	(4.867)	(4.427)			

Notes: Standard deviations reported in parentheses. The extreme right column reports the P-value of an F-test of joint significance of treatment assignment as a predictor of the variable for that row using an OLS regression.

Appendix D: Econometric methods and regression specifications

All regression specifications are fitted using ordinary least squares (OLS). Test scores in each round (baseline or endline) are normalized using the mean and standard deviation of the control group's scores in that round. In each regression, we include indicators for missing covariates and replace missing values with zero to avoid respondents' dropping out of our analysis due to non-response for particular variables.

Primary Schools
For the main results, we regress:

$$Y_{1is} = \beta_0 + \beta_1 T_{Cs} + \beta_2 T_{Ls} + \beta_4 \boldsymbol{Y}_{0oi} + \beta_5 \boldsymbol{Y}_{0wi} + \pi \mathbf{C} + \gamma \mathbf{S} + \delta C + \varepsilon_{is} \quad (1)$$

$$Y_{1is} = \beta_0 + \beta_1 T_{Cs} + \beta_2 T_{Ls} + \beta_3 (T_{Cs} * T_{Ls}) + \beta_4 \boldsymbol{Y}_{0oi} + \beta_5 \boldsymbol{Y}_{0wi} + \pi \mathbf{C} + \gamma \mathbf{S} + \delta C + \varepsilon_{is} \quad (2)$$

Where:

- Y_{1hs} is the endline test score for individual *i*, in school *s*. Regressions are run separately for
  - normalized oral reading (Hindi) exam score;
  - normalized oral math exam score;
  - normalized written Hindi exam total score; and
  - normalized written math exam score.
- T_{Cs} is the school treatment dummy for CCE
- T_{Ls} is the school treatment dummy for LEP
- $\boldsymbol{Y}_{0oi}$are the baseline oral test scores for individual *i*
- $\boldsymbol{Y}_{0wi}$are the baseline written test scores for individual *i*
- **C** is a vector of child characteristics (age, grade, gender)
- **S** is a vector of dummies for block*school campus type*average baseline test score groupings which were used for stratification
- ε_{is} is the individual error term, clustered at the school campus level (the level of randomization).

For local area treatment effects, we restrict regression (1) to the different subgroups of our sample (female vs. male, etc.) in question. Note that these regressions are run with no interaction of T_C*T_L to avoid multiplying interactions.

Upper Primary Schools:
For our upper primary school sample, we regress:

$$Y_{1is} = \beta_0 + \beta_1 T_{Cs} + \beta_2 \boldsymbol{Y}_{0wi} + \pi \mathbf{C} + \gamma \mathbf{S} + \delta C + \varepsilon_{is} \quad (3)$$

Where:

- Y_{1hs} is the endline test score for individual *i*, in school *s*. Regressions are run separately for
  - normalized NCERT Hindi exam score; and
  - normalized NCERT math exam score.
- T_{Cs} is the school treatment dummy for CCE
- $\boldsymbol{Y}_{0wi}$ are the baseline NCERT test scores for individual *i*
- **C** is a vector of child characteristics (age, gender)
- **S** is a vector of dummies for block*school campus type*average baseline test score groupings which were used for stratification
- ε_{is} is the individual error term, clustered at the school campus level (the level of randomization).

Appendix E: Results tables

(Note: For regression specifications, see Annex D.)

Table E1: Main results, primary schools

	Hindi				Math			
	ASER Reading		Written Hindi		ASER Math		Written Math	
CCE	0.00230	-0.00358	0.0283	0.00237	0.00557	-0.00759	0.0138	-0.0152
	(0.0173)	(0.0234)	(0.0207)	(0.0274)	(0.0153)	(0.0233)	(0.0213)	(0.0327)
LEP	0.152***	0.146***	0.135***	0.109***	-0.00581	-0.0189	0.0232	-0.00540
	(0.0167)	(0.0223)	(0.0208)	(0.0285)	(0.0154)	(0.0221)	(0.0222)	(0.0295)
CCE*LEP		0.0115		0.0504		0.0256		0.0564
		(0.0337)		(0.0415)		(0.0307)		(0.0444)
Female	0.100***	0.100***	0.122***	0.122***	-0.101***	-0.102***	-0.00113	-0.00152
	(0.0129)	(0.0129)	(0.0152)	(0.0151)	(0.0129)	(0.0129)	(0.0137)	(0.0137)
Grade at baseline	0.0730***	0.0728***	-0.0245	-0.0250	0.0557***	0.0554***	0.00493	0.00443
	(0.0140)	(0.0140)	(0.0177)	(0.0177)	(0.0139)	(0.0138)	(0.0173)	(0.0173)
Age at endline (months)	0.00335***	0.00335***	0.00264***	0.00263***	-0.000723	-0.000714	0.00160***	0.00158***
	(0.000523)	(0.000523)	(0.000598)	(0.000598)	(0.000496)	(0.000496)	(0.000534)	(0.000534)
Baseline ASER Reading (normalized)	0.621***	0.621***	0.233***	0.232***	0.298***	0.298***	0.0331**	0.0324**
	(0.0118)	(0.0117)	(0.0149)	(0.0148)	(0.0122)	(0.0122)	(0.0154)	(0.0154)
Baseline Written Hindi (normalized)	0.120***	0.120***	0.478***	0.479***	-0.0516***	-0.0520***	0.318***	0.318***
	(0.0171)	(0.0171)	(0.0116)	(0.0116)	(0.0180)	(0.0180)	(0.0118)	(0.0118)
Baseline ASER Math (normalized)	0.161***	0.161***	0.0275*	0.0282*	0.437***	0.436***	0.246***	0.247***
	(0.0131)	(0.0131)	(0.0158)	(0.0158)	(0.0142)	(0.0143)	(0.0158)	(0.0158)
Baseline Written Math (normalized)	-0.0835***	-0.0833***	0.174***	0.173***	0.162***	0.163***	0.311***	0.310***
	(0.0196)	(0.0196)	(0.0125)	(0.0125)	(0.0189)	(0.0189)	(0.0124)	(0.0124)
Constant	0.119*	0.122**	0.249***	0.263***	0.0220	0.0289	0.119*	0.135*
	(0.0611)	(0.0613)	(0.0752)	(0.0755)	(0.0559)	(0.0562)	(0.0697)	(0.0699)
Observations	11,963	11,963	9,204	9,204	11,950	11,950	9,204	9,204
R-squared	0.637	0.637	0.651	0.651	0.652	0.653	0.666	0.666

Notes: Scores at baseline and endline normalized using the mean and SD of the control group in each round. Robust standard errors in parentheses (clustered at school campus level). Regressions include fixed effects for strata used in randomization and dummies for missing covariates (coefficients not reported).
*Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level.

Table E2: Results by gender, primary schools

	ASER Reading		Written Hindi		ASER Math		Written Math	
	Females	Males	Females	Males	Females	Males	Females	Males
CCE	0.0157	-0.0105	0.0466*	0.00683	0.0115	0.000877	0.0354	-0.00505
	(0.0216)	(0.0198)	(0.0252)	(0.0229)	(0.0184)	(0.0192)	(0.0242)	(0.0254)
LEP	0.169***	0.131***	0.146***	0.126***	-0.00562	-0.00728	0.0249	0.0219
	(0.0210)	(0.0192)	(0.0260)	(0.0228)	(0.0189)	(0.0186)	(0.0264)	(0.0254)
Age at endline (months)	0.00434***	0.00237***	0.00326***	0.00210***	0.00149**	0.000112	0.00217***	-0.00103
	(0.000768)	(0.000686)	(0.000839)	(0.000722)	(0.000714)	(0.000699)	(0.000700)	(0.000723)
Grade at baseline	0.120***	0.0257	0.00108	-0.0476**	0.0722***	0.0394**	0.0264	-0.0127
	(0.0212)	(0.0195)	(0.0225)	(0.0240)	(0.0189)	(0.0183)	(0.0219)	(0.0230)
Baseline Written Hindi (normalized)	0.644***	0.597***	0.232***	0.238***	0.311***	0.290***	0.0670***	0.00867
	(0.0156)	(0.0163)	(0.0205)	(0.0213)	(0.0158)	(0.0161)	(0.0207)	(0.0229)
Baseline ASER Reading (normalized)	0.106***	0.133***	0.495***	0.458***	-0.0279	-0.0778***	0.313***	0.328***
	(0.0226)	(0.0238)	(0.0151)	(0.0163)	(0.0237)	(0.0275)	(0.0153)	(0.0165)
Baseline Written Math (normalized)	0.114***	0.203***	0.0512**	0.00360	0.402***	0.461***	0.244***	0.245***
	(0.0194)	(0.0166)	(0.0226)	(0.0213)	(0.0188)	(0.0169)	(0.0218)	(0.0211)
Baseline ASER Math (normalized)	-0.0688***	-0.0940***	0.130***	0.214***	0.184***	0.146***	0.266***	0.344***
	(0.0258)	(0.0250)	(0.0170)	(0.0159)	(0.0253)	(0.0254)	(0.0162)	(0.0161)
Constant	0.174**	0.158**	0.350***	0.266***	-0.0445	-0.0293	0.111	0.105
	(0.0836)	(0.0785)	(0.0997)	(0.0896)	(0.0747)	(0.0822)	(0.0842)	(0.101)
Observations	6,098	5,865	4,668	4,536	6,094	5,856	4,668	4,536
R-squared	0.644	0.633	0.651	0.657	0.656	0.656	0.660	0.680

Notes: Scores at baseline and endline normalized using the mean and SD of the control group in each round. Robust standard errors in parentheses (clustered at school campus level). Regressions include fixed effects for strata used in randomization and dummies for missing covariates (coefficients not reported).
* Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level.

Table E3: Results by in-class heterogeneity, primary schools

	ASER Reading		Written Hindi		ASER Math		Written Math	
VARIABLES	Low Heterogeneity	High Heterogeneity	Low Heterogeneity	High Heterogeneity	Low Heterogeneity	High Heterogeneity	Low Heterogeneity	High Heterogeneity
CCE	-0.00892	0.0299	-0.0126	0.00442	-0.0109	0.0312	-0.0224	0.0135
	(0.0249)	(0.0212)	(0.0297)	(0.0260)	(0.0188)	(0.0213)	(0.0283)	(0.0281)
RE	0.151***	0.156***	0.132***	0.147***	0.00801	-0.0111	0.0182	0.0320
	(0.0230)	(0.0204)	(0.0283)	(0.0249)	(0.0184)	(0.0211)	(0.0316)	(0.0287)
Female	0.0680***	0.129***	0.0582**	0.155***	-0.0764***	-0.118***	-0.0705***	0.00568
	(0.0188)	(0.0164)	(0.0225)	(0.0223)	(0.0166)	(0.0184)	(0.0229)	(0.0202)
Age at endline (months)	-0.00251***	-0.00442***	-0.00497***	-0.00389***	-0.000298	-0.00155**	-0.00265***	-0.00349***
	(0.000797)	(0.000682)	(0.000971)	(0.000840)	(0.000782)	(0.000633)	(0.000824)	(0.000751)
Grade at baseline	0.113***	0.00522	-0.0141	-0.0644**	0.0672***	-0.0121	-0.0233	-0.0470*
	(0.0249)	(0.0184)	(0.0291)	(0.0268)	(0.0216)	(0.0187)	(0.0298)	(0.0278)
Baseline ASER Reading (normalized)	0.602***	0.608***	0.307***	0.342***	0.318***	0.244***	0.143***	0.0937***
	(0.0254)	(0.0136)	(0.0218)	(0.0196)	(0.0215)	(0.0148)	(0.0242)	(0.0212)
Baseline Written Hindi (normalized)	0.118***	0.134***	0.375***	0.340***	-0.0323	-0.0125	0.210***	0.222***
	(0.0376)	(0.0193)	(0.0182)	(0.0170)	(0.0422)	(0.0194)	(0.0184)	(0.0170)
Baseline ASER Math (normalized)	0.226***	0.129***	0.112***	0.0927***	0.541***	0.407***	0.325***	0.327***
	(0.0241)	(0.0153)	(0.0233)	(0.0214)	(0.0300)	(0.0153)	(0.0247)	(0.0203)
Baseline Written Math (normalized)	-0.201***	-0.0237	0.123***	0.0987***	0.131***	0.209***	0.259***	0.230***
	(0.0409)	(0.0212)	(0.0208)	(0.0159)	(0.0503)	(0.0202)	(0.0189)	(0.0165)
Constant	-0.134	0.477***	0.615***	0.655***	-0.0321	0.393***	0.489***	0.624***
	(0.105)	(0.0886)	(0.127)	(0.109)	(0.0960)	(0.0814)	(0.120)	(0.113)
Observations	5,562	6,365	2,988	3,301	5,578	6,335	3,060	3,229
R-squared	0.603	0.621	0.685	0.691	0.568	0.605	0.689	0.690

Notes: For each test, "low/high heterogeneity" indicates that a student was in a class with below-/above-median variance in baseline scores for that particular test. Scores at baseline and endline normalized using the mean and SD of the control group in each round. Robust standard errors in parentheses (clustered at school campus level). Regressions include fixed effects for strata used in randomization and dummies for missing covariates (coefficients not reported).

* Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level.

Table E4: Main results, upper primary schools

	NCERT Hindi	NCERT Math
CCE	0.0225	-0.0409
	(0.0454)	(0.0563)
Female	0.133***	-0.0260
	(0.0325)	(0.0440)
Age at endline (months)	0.00898***	0.00444***
	(0.00111)	(0.00141)
Baseline NCERT Hindi (normalized)	0.581***	0.275***
	(0.0266)	(0.0248)
Baseline NCERT Math (normalized)	0.00299	0.0830***
	(0.0180)	(0.0229)
Constant	1.424***	0.806***
	(0.187)	(0.241)
Observations	2,999	3,000
R-squared	0.420	0.155

Notes: Scores at baseline and endline normalized using the mean and SD of the control group in each round. Robust standard errors in parentheses (clustered at school campus level). Regressions include fixed effects for strata used in randomization and dummies for missing covariates (coefficients not reported).
* Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level.

Appendix F: Scoring criteria

Scoring Criteria for Primary School Tests

For the oral and written tests of basic Hindi and math skills used in the primary school sample, the following scoring and weighting criteria were used.

Category	Competency	Round	Grade in 2011–12 school year	Weighted score
	Oral Exam			
Language	Can read letters	Baseline and Endline	Grades 1–4	1
Language	Can read words—with and without *matra* (vowels)	Baseline and Endline	Grades 1–4	1
Language	Can read a paragraph	Baseline and Endline	Grades 1–4	1
Language	Can read a story (signified by proper inflection, rhythm, and understanding)	Baseline and Endline	Grades 1–4	1
Math	Can identify single digits (1–9, not in sequential order)	Baseline and Endline	Grades 1–4	1
Math	Can identify double digits (11–99, not in sequential order)	Baseline and Endline	Grades 1–4	1
Math	Can successfully do subtraction (double digit without carry-over)	Baseline and Endline	Grades 1–4	1
Math	Can successfully do division (three digits by one digit)	Baseline and Endline	Grades 1–4	1
Math	Present correct amount of currency (2 and 3-digits)	Baseline and Endline	Grades 1–4	1
Math	Word problems with (subtraction and division)	Baseline and Endline	Grades 1–4	1
Language	**Baseline Total**			**4 (scaled to 6)**
Math	**Baseline Total**			**6**
Language	Comprehension of paragraph	Endline	Grades 1–4	1
Language	Comprehension of story	Endline	Grades 1–4	1
Language	**Endline Total**			**6**
Math	**Endline Total**			**6**
Category	**Competency**	**Round**	**Grade in 2011–12 school year**	**Weighted score**
	Written Exam			
Language	Write (spoken) letter	Baseline and Endline	Grades 2–4	1
Language	Write (spoken) word	Baseline and Endline	Grades 2–4	1
Language	Match correct word (from several options) to picture	Baseline and Endline	Grades 2–4	1
Language	Word completion (based on picture)	Baseline and Endline	Grades 2–4	1
Language	Write word (based on picture)	Baseline and Endline	Grades 2–4	1

Category	Competency	Round	Grade in 2011–12 school year	Weighted score
Language	Select correct antonym (from several options) of written word	Baseline and Endline	Grades 2–4	1
Language	Sentence completion (choose word)	Baseline and Endline	Grades 2–4	1
Language	Sentence completion (write word)	Baseline and Endline	Grades 2–4	1
Language	Write sentence describing picture	Baseline and Endline	Grades 2–4	1
Language	Written answer signifying comprehension of paragraph	Baseline and Endline	Grades 2–4	1
Language	Written answer signifying comprehension of story	Baseline and Endline	Grades 2–4	1
Math	Count objects and choose correct number	Baseline and Endline	Grades 2–4	1
Math	Complete numerical sequence (1 digit)	Baseline and Endline	Grades 2–4	1
Math	Complete numerical sequence (2 digits)	Baseline and Endline	Grades 2–4	1
Math	Identify written long form of number	Baseline and Endline	Grades 2–4	1
Math	Write out long form of given number	Baseline and Endline	Grades 2–4	1
Math	Identify largest number from numerals	Baseline and Endline	Grades 2–4	1
Math	Identify time from illustration of clock face	Baseline and Endline	Grades 2–4	1
Math	Addition (1–3 digits, with/without carrying)	Baseline and Endline	Grades 2–4	1
Math	Subtraction (1–3 digits, with/without carrying)	Baseline and Endline	Grades 2–4	1
Math	Addition (four digits with carrying)	Baseline and Endline	Grades 2–4	1
Math	Subtraction (four digits with carrying)	Baseline and Endline	Grades 2–4	1
Math	Multiplication (1–2 digits)	Baseline and Endline	Grades 2–4	1
Math	Word problems (subtraction and multiplication)	Baseline and Endline	Grades 2–4	1
Math	Division (3-digit dividend, 1-digit divisor)	Baseline and Endline	Grades 2–4	1
Language	**Total**	**Baseline and Endline**	**Grades 2–4**	**11 (scaled to 12.5)**
Math	**Total**	**Baseline and Endline**	**Grades 2–4**	**14 (scaled to 12.5)**

References

ASER Centre, 2011a. *Annual Status of Education Report 2011*. New Delhi: Pratham.

ASER Centre, 2011b. *ASER District Performance Tables 2011.* Available at: http://images2.asercentre.org/aserreports/aser_2011_district_performance_tables_revised.pdf.

ASER Centre, 2013. *Annual Status of Education Report 2012*. New Delhi: Pratham.

Banerjee, Abhijit V, Cole Shawn, Duflo, Esther, and Linden, Leigh, 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, 122(3), pp.1235–64.

Banerjee, Abhijit V, Banerji, Rukmini, Duflo, Esther, Glennerster, Rachel, and Khemani, Stuti, 2010. Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India, American Economic Journal: Economic Policy, 2(1), pp.1–30.

Banerjee, Abhijit V, Banerji, Rukmini, Duflo, Esther, and Walton, Michael, 2011. "Effective Pedagogies and a Resistant Education System: Experimental Evidence on Interventions to Improve Basic Skills in Rural India," Working paper.

Black, Paul and Wiliam, Dylan, 2004. The Formative Purpose: Assessment Must First Promote Learning, *Yearbook of the Society for the Study of Education*, 103(2), pp.20–50.

Black, Paul and Wiliam, Dylan, 2009. Developing the Theory of Formative Assessment, *Educational Assessment, Evaluation and Accountability*, 21(1), pp.5–31.

Bolyard, KJ, 2003. "Linking Continuous Assessment and Teacher Development: Evaluating a Model of Continuous Assessment for Primary Schools in Malawi," Educational Quality Improvement Program, USAID.

Census of India, 2011. Houselisting and Housing Census Data Highlights. New Delhi: Office of the Registrar General and Census Commissioner, India.

Central Board of Secondary Education (India), 2009. Quarterly Bulletin of the Central Board of Secondary Education, 48(1).

Duflo, Annie and Kiessel Jessical, 2012. "Every Child Can, Every Child Counts: An evaluation of the Teacher Community Assistant Initiative Pilot Programme in Ghana," Working paper.

Duflo, Esther, Dupas, Pascaline, and Kremer, Michael, 2011. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, *American Economic Review*, 101(5), pp.1739–74.

Government of India, 2009. The Right of Children to Free and Compulsory Education Act 2009. Gazette of India, 39(27 August).

International Institute for Population Sciences (IIPS) and Macro International, 2008. National Family Health Survey (NFHS-3), India, 2005–06. Mumbai: IIPS.

Kamangira, YT, 2003. "Feasibility of a Large Scale Implementation of Continuous Assessment as a Stimulus for Teacher Development in Malawi," Improving Educational Quality (IEQ) Project.

Kapambwe, William M, 2010. The Implementation of School Based Continuous Assessment (CA) in Zambia, *Educational Research and Reviews*, 5(3), pp.99–107.

National Council of Educational Research and Training (India), 2005. *National Curriculum Framework.*

National Council of Educational Research and Training (India), 2013. Available at: http://www.ncert.nic.in [Accessed 1 December 2013].

National Institute of Rural Development (India), 2009. "A Note on the Backward Regions Grant Fund Programme," Ministry of Panchayati Raj, September.

National University of Educational Planning and Administration (India), 2013. *Elementary Education in India: Where Do We Stand?* New Delhi: NUEPA.

Reserve Bank of India, 2011. *Handbook of Statistics on Indian Economy 2010–11*. New Delhi: RBI.

Publications in the 3ie Impact Evaluation Report Series

The following reports are available from http://www.3ieimpact.org/evidence-hub/publications/3ie-impact-evaluation-repository

The promise of preschool in Africa: A randomised impact evaluation of early childhood development in rural Mozambique, *3ie Impact Evaluation Report 1*. Martinez, S, Naudeau, S and Pereira, V (2012)

A rapid assessment randomised-controlled trial of improved cookstoves in rural Ghana, *3ie Impact Evaluation Report 2*. Burwen, J and Levine, DI (2012)

The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone, *3ie Impact Evaluation Report 3*. Casey, K, Glennerster, R and Miguel, E (2013)

Does marginal cost pricing of electricity affect groundwater pumping behaviour of farmers? Evidence from India, 3ie Impact Evaluation Report 4. Meenakshi, JV, Banerji, A, Mukherji, A and Gupta, A (2013)

Impact evaluation of the non-contributory social pension programme 70 y más in Mexico, *3ie Impact Evaluation Report 5.* Rodríguez, A, Espinoza, B, Tamayo, K, Pereda, P, Góngora, V, Tagliaferro, G and Solís, M (2014)

The impact of daycare on maternal labour supply and child development in Mexico, *3ie Impact Evaluation Report 6*. Angeles, G, Gadsden, P, Galiani, S, Gertler, P, Herrera, A, Kariger, P and Seira, E (2014)

Social and economic impacts of Tuungane: final report on the effects of a community-driven reconstruction programme in the Democratic Republic of Congo, *3ie Impact Evaluation Report 7*. Humphreys, M, Sanchez de la Sierra, R and van der Windt, P (2013)

Paying for performance in China's battle against anaemia, *3ie Impact Evaluation Report 8*. Zhang, L, Rozelle, S and Shi, Y (2013)

No margin, no mission? Evaluating the role of incentives in the distribution of public goods in Zambia, *3ie Impact Evaluation Report 9*. Ashraf, N, Bandiera, O and Jack, K (2013)

Truth-telling by third-party audits and the response of polluting firms: Experimental evidence from India, 3ie Impact Evaluation Report 10. Duflo, E, Greenstone, M, Pande, R and Ryan, N (2013)

An impact evaluation of information disclosure on elected representatives' performance: evidence from rural and urban India, 3ie Impact Evaluation Report 11. Banerjee, A, Duflo, E, Imbert, C, Pande, R, Walton, M and Mahapatra, B (2014)

Targeting the poor: evidence from a field experiment in Indonesia, 3ie Impact Evaluation Report 12. Atlas, V, Banerjee, A, Hanna, R, Olken, B, Wai-poi, M and Purnamasari, R (2014)

Scaling up male circumcision service provision: results from a randomised evaluation in Malawi, 3ie Impact Evaluation Report 13. Thornton, R, Chinkhumba, J, Godlonton, S and Pierotti, R (2014)

Providing collateral and improving product market access for smallholder farmers: a randomised evaluation of inventory credit in Sierra Leone, 3ie Impact Evaluation Report 14. Casaburi, L, Glennerster, R, Suri, T and Kamara, S (2014)

A youth wage subsidy experiment for South Africa, 3ie Impact Evaluation Report 15. Levinsohn, J, Rankin, N, Roberts, G and Schöer, V (2014)

The impact of mother literacy and participation programmes on child learning: evidence from a randomised evaluation in India, 3ie Impact Evaluation Report 16. Banerji, R, Berry, J and Shortland, M (2014)

Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua, *3ie Impact Evaluation Report 17*. Barham, T, Macours, K, Maluccio, JA, Regalia, F, Aguilera, V and Moncada, ME (2014)

A randomised evaluation of the effects of an agricultural insurance programme on rural households' behaviour: evidence from China, 3ie Impact Evaluation Report 19. Cai, J, de Janvry, A and Sadoulet, E (2014)

A wide angle view of learning: evaluation of the CCE and LEP programmes in Haryana, 3ie Impact Evaluation Report 22. Duflo, E, Berry, J, Mukerji, S and Shotland, M (2015)

National educational surveys in India have consistently shown that the vast majority of students fail to attain grade-level competencies at the end of five years of primary schooling.

This evaluation compares the impact of the Continuous and Comprehensive Evaluation (CCE) scheme and the Learning Enhancement Programme (LEP), as well as a combination of the two approaches, on student achievement in the state of Haryana in India. While the CCE aims to provide teachers and students with broad-based and frequent feedback on performance, the LEP offers tools and allocates time within the school day to enable teachers to focus their teaching at each child's competency level.

The impact evaluation finds that students in CCE schools did not perform significantly better than their peers in control schools. In contrast, the LEP had a large, positive and statistically significant effect on students' basic reading abilities. Combining CCE and LEP had no significant effect on student test scores relative to the LEP alone.

Impact Evaluation Series

International Initiative for Impact Evaluation
202-203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org