

# **The Andhra Pradesh Randomized Evaluation Study**

## **Summary**

## Table of Contents

Section 1: Motivation.....	1
Section 2: The Partnership.....	3
Section 3: The Design of the Research Program .....	5
Section 4: Results and Findings .....	9
Section 5: Summary and Conclusions .....	17
References.....	19

This summary report presents the overall results from what is nearly a ten-year long research program entitled the Andhra Pradesh Randomized Evaluation Study (APRESt). APRESt consists of two major strands of research – on the relative effectiveness of schooling inputs and teachers incentives, and on the impacts of school choice<sup>1</sup>, towards improving the learning outcomes of students. The research program analyzes the impacts of these various policy options through a series of large-scale randomized evaluations. The rest of the summary report is organized as follows: Section 1 and 2 focuses on the genesis of the research program, the partnership built for the program, and the roles played by each partner. Section 3 describes the project interventions, specifically the experimental design, and Section 4 summarizes the overall results. Section 5 presents broad conclusions and recommendations. The report is meant to be read as an overview or an executive summary of the results obtained from this research program. The appendix includes all technical papers produced till date.

---

<sup>1</sup> These results are still being finalized and are not presented in this summary.

## Section 1: Motivation

The responsibility for the delivery of education services is shared between the Center and States in India. Successive governments, both at the center and in state capitals, have made efforts to improve access, equity and quality to basic education. The last decade has witnessed the implementation of large government programs, far-reaching policy developments and innovative and effective solutions being developed, piloted and implemented. These include *inter alia*:

- The Sarva Shiksha Abhiyaan (SSA) program, a centrally sponsored scheme supports a partnership between the federal and state level actors and was launched as a nation-wide effort to ensure that all children of the appropriate age were enrolled in basic schooling. This program has rapidly expanded the schooling system in the country and has led to enormous increases in student enrolment, particularly in the poorest states of the country.
- In 2009-10, the Government of India passed an extremely important legislation popularly known as the Right to Education Act (RTE). The RTE aims to make free and compulsory education a fundamental right for all children in the country.
- In 2001, in response to public interest litigation, the Supreme Court ordered the states to provide cooked or prepared midday meals in every government and government assisted school. This is credited with attracting some of the poorest kids into schools across the country. To meet this mandate, the central government in collaboration with the state governments, launched another centrally sponsored scheme known as the Midday Meal Scheme (MDMS).
- States also have the flexibility to introduce and implement innovative schemes. Perhaps the most notable program in recent years being Tamil Nadu's Activity Based Learning (ABL) Program. This aims to alter the manner in which the teachers engage with students in the classroom, and have children learning at their own pace and by engaging in activities rather than in the more traditional lecture mode.

Though inputs to school education have increased significantly over the last fifteen years, there are growing concerns that these are not necessarily being translated into outcomes, and particularly into learning outcomes.

Several recent assessments demonstrate this concern in translating increasing inputs into meaningful outcomes. In a recent round of the Program for International Student Assessment (PISA), India ranked 73 out of 74 nations, with only Kyrgyzstan performing more poorly. The PISA results are even more sobering when one considers the fact that the two states which participated in the PISA are considered national front runners in education - Himachal Pradesh and Tamil Nadu.

Beyond these international assessments, there are numerous assessments within India that reiterate this message on learning outcomes. For example, Pratham's 2012 Annual Survey of Education Report (ASER) finds very poor levels of learning in Reading and Mathematics. In reading, ASER finds that over half the children in Grade 5 were unable to read a Grade 2 text and more than 60% of kids in Grade 3 were unable to read a Grade 1 text. Similarly, they find that almost half of Grade 5 students were unable to perform a two digit subtraction problem and three-quarters were unable to perform division in a test of basic numeracy skills. Though there may be concerns with precision of ASER assessments, the tools

themselves are simple enough to point towards a very serious problem in quality of the schooling that children receive in typical schools in across the country.

## Section 2: The Partnership

The art and science of delivery refers to an approach by which the creative potential within and sense of partnership across individuals, societies, other stakeholders and governments can be harnessed, and combined with rigorous efforts to understand what works and what does not work in ensuring results - results for those who need it the most, results in a manner that is at scale, effective in outcomes, efficient in comparison to other alternatives, and sustainable. The implementation of the APRESt is a strong example of developing and demonstrating the art and science behind the delivery of essential schooling services.

APRESt emerged from a common desire across key stakeholders to improve learning outcomes in government schools and the recognition that teacher motivation and effort was an issue that needed to be studied more deeply. There were several reasons behind this decision. Firstly, two influential reports identified teacher motivation as an issue across India: (i) a World Bank-Harvard survey on teacher and health provider absence and (ii) very important work done earlier and popularly referred to as the PROBE Report. Secondly, since nearly 90 percent of primary education spending went towards teacher salaries, it was evident that this was an area that needed greater attention and scrutiny. While there is a broad understanding that achieving primary education goals requires a set of inputs be brought to bear on the issue, it is not clear how, which and in what quantities these inputs should be brought together in a classroom setting to produce the desired results. If the vast majority of those resources are for teacher salaries, this leaves very little for child-level learning materials such as notebooks, exercise books, and writing materials and hence has serious implications for public resource allocations and use. Thirdly, since the Indian government was planning to continue raising primary education budgets to meet the schooling targets, there was a serious need to understand the most efficient way to spend these scarce public resources. Finally, there was broad recognition that salaries paid to any government employee were largely disconnected from any measures of performance and that this applied to civil service teachers as well and that there might be potential gains from experimenting with interventions that aligned the incentives of teachers to those of the government or state in terms of improving learning outcomes. At the same time, there was a broad recognition that there might be policy options, as yet unexplored and untested in the Indian context, which could raise learning levels in classrooms.

The issues above brought together policy makers, teachers and teacher union members, education experts, experts at student assessments, academics and development practitioners. There were four main parties to this partnership: the Government of the State of Andhra Pradesh (GOAP), the Azim Premji Foundation (APF), UKAID and the World Bank (the Bank). The Government of India – through its Ministries of Human Resources Development (MHRD) and Finance too supported the program. A working group was formed consisting of representatives of the above organizations.

The Government of Andhra Pradesh: It was agreed that the GOAP would play the lead role in terms of oversight and in creating an authorizing environment under which APRESt would be implemented. This decision was particularly important given concerns of legitimacy of the work, but also because the GOAP was a key stakeholder with a keen interest in the issue of teacher motivation. They partnered with the Bank to ensure that this would be a learning exercise and helped finalize key design elements. Of central importance, the GOAP agreed to randomize treatments across schools thereby enabling an experimental

design to be put into place. This experimental design ensured that the research study findings are of highest rigor, which gives all stakeholders great confidence in the validity of its results. The GOAP went a step further and also provided financial and in-kind support for this research program and entered into a Memorandum of Understanding with the APF through which approximately USD 500,000 was allocated for this research. Furthermore, the GOAP supported the intervention by allocating contract teachers to a select set of 100 schools determined through a randomization algorithm. Finally, the GOAP ensured that transfers into and out of the school identified for the program would be frozen for a period of two years.

The Azim Premji Foundation: The APF is the largest education oriented NGO in India and is one of the pioneers of NGO-Government collaboration. Many NGOs in the education sector in India operate in a parallel space to that occupied by the Government and see themselves as substitutes to the government. The APF's approach has always been to acknowledge the important role that government plays in the delivery of education services across the country and the need to partner with the government to improve the opportunities for 75-80% of the children in the school system in India who rely on access and quality in government schools. The APF does this through many different interventions including school level incentives through the learning guarantee program, digital content for schools, student and teacher assessments and supporting teacher training. More recently, the APF has set up a university in Bangalore to support the training of teachers and research in the education sphere. For all these reasons, the APF was an obvious choice to help implement or support any program on the ground aimed at boosting teacher motivation and performance in Andhra Pradesh. The Foundation had credibility with the government and more importantly, credibility with the teachers and teacher unions. Given the nature of the this was an extremely important consideration.

United Kingdom Agency for International Development (formerly DFID): While the roles played by GOAP and APF were central to the success of this program, the role of UKAID was nothing short of critical. Taxpayer money from the United Kingdom financed most of this intervention and placed a serious responsibility on team to ensure the delivery of high quality analytical products. UKAID financing supported all aspects of the program from design, implementation, analysis and report writing. In addition, the program has also supported dissemination events. UKAID financing was also instrumental in helping leverage program financing from other sources.

The World Bank: The role of the Bank in this activity has been a central one – in many ways the glue that held everything together - from the concept stage, to bringing this diverse set of actors around to a common platform, technical support and administrative and institutional oversight. The World Bank used its convening power to ensure institutional continuity and provided technical guidance at every stage

### Section 3: The Design of the Research Program

This working group adopted three key guiding principles in establishing the research parameters:

- (i) the study should push both academic and policy frontiers to the extent possible;
- (ii) interventions should be based on evidence and not ideology, and no intervention should be ruled out before a thorough discussion of the pros and cons; and
- (iii) to the extent possible the evaluation should be conducted over several years to ensure credibility and sustainability of the results.

Given the above key guiding principles and based on the factors that led to establishment of this research program, the working group agreed to experiment with teacher incentives in an effort to motivate their effective participation in the schooling system.

The use of incentives as a way of improving performance in the schooling system was not a new concept in India. Teacher recognition programs have been in place for many years in the country and every year a handful of teachers are recognized publicly either in their own districts or at state or national functions. In addition, organizations such as the APF have run their own programs in partnership with state governments for many years. For example, the APF has implemented a program known as the Learning Guarantee Program (LGP) that provided cash incentives to government schools which self-selected into the program and for achieving specified learning levels. Both in-kind incentive programs of state governments and the LGP are typically conducted as tournaments – where the best performing teacher (or set of teachers) and the best performing school (or set of schools) receive the incentives. While the measures of performance under the LGP were very clear and explicit (see Barnhardt 2007 for further details), in most cases the rules applied to determine the best teacher in the district or state by the government have been vague. During focus group interviews with teachers, three clear issues emerged in terms of the available government led teacher recognition schemes: (i) selection criteria were not clear, (ii) teachers often felt that subjective considerations played a role in determining the “best teacher” awardee, and (iii) while in-kind recognitions were indeed valuable, the teacher’s themselves repeatedly stated that they “*would not be opposed to receiving cash bonuses for performance*”. Their main concerns seemed to focus on the fact that explicit, clear, and objective rules of the game were needed and that there had to be an honest broker and objective means of measuring or assessing teacher performance in place for such a scheme to work.

As individual, in-kind incentive programs, and school level incentive programs had been used in the past in the Indian context, and based on focus group interviews with teachers, the working group decided to focus on *cash incentives to teachers* as one of the interventions. As this was the first time to our knowledge that such a program would have been put in place in government schools, and for government school teachers in India, the team believed it was important to ensure that the findings could be generalized across states, and that there should be a meaningful comparison of such a policy option against the more typical investments made in government schools in India which include *inter alia*: (i) infrastructure inputs, (ii) teacher and teaching inputs, and (iii) teaching and learning materials for students.

Therefore, the working group agreed to evaluate the relative returns to additional spending on typical schooling inputs on the one hand, against a policy of trying to improve student level outcomes by directly

incentivizing teachers through the potential for receiving cash benefits for improved class performance on the other hand.

### **The Inputs:**

The above three input choices were considered. The working group quickly ruled out improvements to infrastructure for two main reasons. Firstly, we were not sure that the necessary improvements to infrastructure could be completed in the short period of the study<sup>2</sup>. Secondly, and perhaps more importantly, the team was concerned that any returns to infrastructure spending in terms of improved learning outcomes may not be witnessed during the study period. So, the working group agreed to focus on teacher and teaching inputs and student teaching and learning material.

To determine the specific nature of the inputs, the working group approached this by asking how an additional rupee available for spending on primary schooling could be best spent. For example, on an additional full-time or contract teacher or on teacher training programs, etc. Interviews with teacher groups and with administrators and other stakeholders suggested that teachers typically received both pre- and in-service training, and at least anecdotally these seemed to have little sustained impact on classroom processes or teacher behavior necessary to improve learning outcomes. So, the working group considered experimenting by looking directly at the returns to having an additional full time teacher versus an additional contract teacher in the classrooms. A direct comparison on the relative effectiveness of civil service teachers and contract teachers would have helped provide evidence on issues of enormous policy relevance – entry requirements, terms and conditions of employment, tenure and career path. However, the GOAP noted administrative and bureaucratic constraints in allocating regular teachers across a randomly selected set of schools, although they noted that it would be relatively easy to assign contract teachers to a randomly selected set of schools<sup>3</sup>. The second input consisted of a block grant to schools to support the purchase of child specific teaching learning materials. Given expenditure levels and patterns of spending and the fact that over 90 percent of public expenditures went to cover teacher salaries and pension liabilities, the working group felt that any additional spending on child-specific educational inputs (such as notebooks, exercise books, writing/coloring materials, etc.) would presumably have the greatest impact<sup>4</sup>.

### **The Incentives**

As stated above, the working group agreed based on feedback from focus group discussions with teachers and on the fact that incentives schemes were regularly used in the Indian context to try and motivate teacher performance, decided to work on teachers incentives but with a twist – to focus on individual and group monetary incentives to teachers conditional upon improved performance on independently conducted standardized assessments. The use of incentives for teachers is really at the forefront of policy debate and discussions in many countries across the world today. APRES bears relevance for teacher

---

<sup>2</sup> At the time the program was initiated, the program was expected to continue for a 2 year period.

<sup>3</sup> Procedures followed hiring of contract teachers are provided in detail in the paper entitled “ “ in the Annex.

<sup>4</sup> The procedures employed under the research program to monitor how block grants were spent and the nature of the items procured are provided in detail in the paper entitled “ “ in the Annex.

policies in numerous developing countries, but also contributes to a wider debate on teacher incentives in education policy even in countries that are further along the development path.

Even though the working group agreed to experiment with financial incentives, it was not without debate. The group felt that restructuring teacher pay to incorporate an incentive conditional upon improved student performance was a reasonable idea, but differed strongly on both ideological and technical grounds, on the nature of the design and the types of incentives. Firstly, there is a widespread belief that in professions such as teaching, medicine, etc., intrinsic motivation is perhaps stronger driver of performance and enhanced effort, and that the use of external incentives could negatively impact on this intrinsic motivation. Secondly, the GOAP was concerned that individual and group incentives could lead to conflicts among teachers who receive and those who may not receive bonus payments within a given school. However, the Key Guiding Principles adopted earlier helped the working group to navigate through such debates. The working group noted that *ex ante* that group incentives may be more effective than individual incentives especially depending on the size of the school with smaller schools being able to monitor individual effort levels better and ensure that other teachers are not free-riding and in larger schools where monitoring becomes more difficult and costly, individual incentives may prove to be more effective. , The group agreed that it was worth experimenting with both group and individual incentives.

Finally, the working group focused on the nature of the experiment. Government programs are typically universal in nature, or they are targeted towards particular segments of the population (e.g., the poorest, people belonging to particular caste or ethnicity, or other sub-groups), or are based on self-selection into programs (such as in the case in a large number of anti-poverty programs). Each of these different approaches in typical programs makes evaluation difficult as it is often difficult to develop a statistically-valid comparison group, identical in all respects to program participants, except that they did not participate in the program of interest. Therefore, to ensure that any program of this nature could be evaluated and causal impacts determined, the technical team proposed a randomized control trial or RCT. Although widely used in clinical trials for new drugs, RCTs have now made inroads into the evaluation of social programs. By matching schools on observable characteristics *ex-ante*, and then randomly assigning some into treatment(s) group(s) and others into a control group allows for a rigorous way of learning about the effectiveness of the program(s). Since the program is randomly assigned to a subset of the potential recipients, the remaining potential recipients provide a perfect control group as there is no selection bias in who received the treatment and who did not. Also, other extraneous factors should affect both groups similarly on average and their effect can be netted out by calculating a “difference in differences” estimate. The methodology of randomized evaluation is considered the “gold standard” in evaluating the causal impact of programs<sup>5</sup>.

---

<sup>5</sup> While random allocation of the treatment ensures internal validity, typically concerns about the generalizability of the results outside of the study sample tend to be difficult. In this case as well this may be true. However, there are some mitigating factors and design innovations that we have included. Firstly, the indicators of interest – teacher absence and teaching activity, etc. are very similar to all India averages. Secondly, during sampling we have made an effort to ensure that the five districts from which the schools are chosen cover all three socio-cultural regions of the state.

## Program Design

A schematic of the research design is provided below. Each individual orange cell represents a unique “treatment”. The yellow cell represents the control group<sup>6</sup>. The relative effectiveness of each treatment can be compared to the control group and also relative to each other. Random assignment to a subset of the universe of possible recipients provides for a perfect control group as there is no selection bias in who received the treatment and who did not. Furthermore, if there are other factors that may have affected all groups during the course of the program, these extraneous factors are likely to impact on groups similarly and thus it is possible to net out their effects by using a “difference in difference” estimate.

		INCENTIVES (Conditional on Improvement in Student Learning)		
		NONE	GROUP MONETARY	INDIVIDUAL MONETARY
INPUTS (Unconditional)	NONE	CONTROL GROUP (100 Schools)	100 Schools	100 Schools
	EXTRA TEACHER	100 Schools		
	EXTRA BLOCK GRANT	100 Schools		

APRESt experiments were chosen purposefully to compare the relative effectiveness of input- and incentive- based policies for improving education quality. Traditionally, Indian education policy has taken an input-focused approach. Incentive-based policies are a relatively new phenomena, but increasingly popular in many parts of the world. The table below summarizes the motivation behind each experiment and briefly describes each.

	Motivation	Description
<b>Diagnostic Feedback</b>	<ul style="list-style-type: none"> <li>One reason learning levels may be low is teachers don't know how to help students</li> <li>Can better information help?</li> </ul>	<ul style="list-style-type: none"> <li>Existing teachers provided with detailed feedback on students and subject to low-stakes monitoring</li> </ul>
<b>Block Grants</b>	<ul style="list-style-type: none"> <li>Significant amounts of money committed under RTE.</li> <li>What is the effectiveness of such spending?</li> </ul>	<ul style="list-style-type: none"> <li>Schools provided cash grants for student inputs</li> </ul>
<b>Contract Teachers</b>	<ul style="list-style-type: none"> <li>Use of contract teachers is widespread, but highly controversial</li> <li>Are contract teachers effective?</li> </ul>	<ul style="list-style-type: none"> <li>Schools provided with additional teacher (on contract)</li> </ul>
<b>Performance Pay x2</b>	<ul style="list-style-type: none"> <li>Teacher salaries are the largest component of education spending, but a poor predictor of outcomes</li> <li>Can linking pay to performance improve outcomes?</li> </ul>	<ul style="list-style-type: none"> <li>Teachers eligible for bonuses based on improved student performance (either in own class or whole school)</li> </ul>

<sup>6</sup> Control schools also received feedback on student performance, guidance to teachers, and the additional monitoring of classrooms due to the above activities. Our experimental evidence extends to reviewing the impact of a program that provided low-stakes diagnostic tests, monitoring of classroom processes, and feedback to teachers on the performance of their children. While we find teachers in treatment schools exerting more effort when observed in the classroom, we also find that students in these schools do no better on independently-administered tests than students in comparison schools which do not receive the program. This suggests that though teachers in the program schools worked harder while being observed, there was no impact of the feedback and monitoring on student learning outcomes. Our study therefore suggests that enhanced monitoring alone, with neither punitive actions for poor performance or rewards for good performance, are unlikely to lead to improved learning in government, primary school classrooms at the moment.

## Section 4: Results and Findings

In this section we classify results into three different sections based on the inputs and incentives studied in this experiment. The first part looks at the role of contract teachers and the impact of the same on improving learning outcomes. The second part looks more closely at the role of schooling inputs in the form of block grants for teaching learning material. The final part reviews the evidence from both individual and group incentives for teachers as a function of increased student test scores.

### Contract Teachers Treatment

While there are many disputed claims in education research, most would agree that a good teacher can make all the difference to improving student learning outcomes in their classrooms<sup>7</sup>. However, the question of what makes a good teacher is highly debated. A growing body of education research acknowledges the importance of teacher quality on demonstrated student achievements, but very little is known about which measurable characteristics of teachers truly influence classroom outcomes. Studies have investigated the impact of licensing, licensure test scores, higher qualifications (such as bachelors, masters or PhDs), and experience<sup>8</sup>. The absence study found a perverse effect that teachers who were better paid or more senior (since pay and seniority are highly correlated in the Indian government school context) and more educated gained greater satisfaction from being absent.

The typical rural government primary school is small, with an average enrollment of about 80 to 100 students. Multi-grade teaching is the norm, and schools typically have 2 to 3 teachers covering grades one through five. A teacher tends to teach all the subjects for one grade and typically teaches in more than one grade. Civil service or regular teachers (RTs) are state employees, have tenure till official retirement age, and have a pensionable job with benefits. These teachers are typically selected through a teacher selection commission or similar means and tend to be more educated (than contract teachers) and have formal teacher training degrees. Contract teachers (CTs) on the other hand are hired at the school level, do not have tenure and their contracts are renewed annually conditional upon performance, and are not considered as state employees. They are typically less qualified than RTs with high school or first degree completion, but most do not have formal teacher training. They also do not have any benefits. The average RT salary plus benefits was approximately Rs. 10000 in 2006, or about five times the compensation package received by CTs in the state.

The experiment in APRES<sub>t</sub> mimics the process by which contract teachers are typically hired in the state. Schools apply to the local district education administration and seek permission for hiring a contract

---

<sup>7</sup> Rigorous studies mostly from the United States suggests that having a good teacher for 3-5 years is enough to eliminate the achievement gaps between whites, blacks or Hispanics. For more on similar results see Rivkin, Hanushek, and Kain (2005) and Kane and Staiger (2008).

<sup>8</sup> On licensing and license test scores please refer to Goldhaber and Brewer (2000), Angrist and Guryan (2003), Buddin and Zamarro (2008), and Aaronson et al (2007); on qualifications please refer to Clotfelter et al. (2006, 2007), Rockoff (2004) and Rockoff and Staiger (2010) and for experience please refer to Clotfelter et al (2006), Rivkin, Hanushek, and Kain (2005) and Ladd (2008)

teacher on the basis of student enrollment strength at the start of the year<sup>9</sup>. Contract teacher contracts in AP typically run for a period of 10 month beyond which their contracts are renewed if there is continued or perceived need. However, in practice, once a position has been sanctioned these were typically continued from year to year<sup>10</sup>. Under the APRESt, schools were initially identified through a randomization process and then were informed through a letter from the district administration that they were authorized to hire an additional contract teacher for the year. This hiring came after all the transfers and all the requests for contract teachers by these schools had been completed for that year. Therefore, the additional contract teacher truly represents an additional teaching staff to the schools and not simply the filling of an existing vacancy or assistance to meet unexpected higher enrollments and hence higher than norm school level PTRs. Once allocations to schools were made, the decision on how and where to use them was strictly left to schools to decide. The state administration also played a crucial role in protecting the schools for the first two years of the research program by ensuring that teachers were not transferred in or out of the sample schools. Four-fifths of the schools initiated the hiring procedures for contract teachers within a week of receiving the authorization and based their hiring on qualifications, experience, and distance from school.

From our sample of RTs and CTs we find the following characteristics:

- (i) RTs are overwhelmingly male compared to CTs (66% to 28%)
- (ii) on an average older by about 15 years
- (iii) twice as likely to have a college degree (85% to 47%)
- (iv) 7 times as likely to have a formal training or education certificate (99% to 13%)
- (v) almost twice as likely to have received training in the last year (92% to 60%)
- (vi) very unlikely to be from the same village in which the school resides (9% to 82%)
- (vii) having to travel 12 times as much as a CT to reach school (12 km to 0.84 km)
- (viii) earn about 9 times as much as a typical CT in our sample (INR 9000 per month to INR 1000 per month).

## Results

The Contract Teacher treatment finds the following:

- (i) At the end of two years of the program, students in schools with an extra contract teacher perform significantly better than children in control schools by 0.15 and 0.13 standard deviations in Math and Telugu respectively.
- (ii) The program has differential effects on different groups of students. We find that students in schools that had poorer infrastructure and were more remote seemed to benefit more by having an additional contract teacher. Secondly, we find that the impact of having an additional contract teacher is highest in Grade 1 averaging 0.23 and 0.25 Standard Deviation increases in Math and Telugu scores, with the treatment declining for higher grades.
- (iii) We find that on key metrics, regular teachers and contract teachers behave differently. RTs are more likely to be absent from school compared to CTs (27% to 16%), and conditional on being in school CTs are more likely to be engaged in teaching activities compared to RTs (49% to 43%). Similar results

---

<sup>9</sup> However, contract teachers can also be hired to fill in vacancies that may arise due to retirement or death or transfers.

<sup>10</sup> Govinda and Yazali (2004) find that almost all para teacher contracts were being renewed year on year.

have been replicated across other studies and therefore suggest that CTs do behave differently than RTs and in a manner more beneficial to enhancing student outcomes.

(iv) By comparing absence rates and teaching activity rates for RTs in schools with and without an extra contract teacher, we find that programs schools do have higher rates of regular teacher absence (28% to 25%) and lower rates of teaching activity by regular teachers (45% to 41%). This suggests that the placement of a CT might induce RTs to shirk even more. Therefore, our estimate of the impact of an additional contract teacher is actually an underestimate, since this is the composite effect of an additional contract teacher minus the negative effect on RT performance given the induction of a CT into the school.

### Policy Directions

There is a tendency for all stakeholders to try and focus on observable teacher characteristics as a means of ensuring a high quality teacher in the classroom. Using these metrics, the RTE has put into effect procedures to eliminate contract teachers in the country and ensure that all teaching staff in government run schools are full time, civil service teachers. As noted earlier, there is little rigorous evidence to support such a policy. However, as shown through this study and others, these measures do not necessarily correlate with success in the classroom. The RTE therefore will have significant fiscal implications for the India, and if existing evidence holds, may not lead to the desired improvements in education quality or in learning outcomes<sup>11</sup>. While the policy decision on contract teachers in the RTE has been taken, these will have no direct implication. However, given the findings of this study it is perhaps important to review the rigorous evidence that does exist to support a review of the policy on contract teachers which finds that contract teachers are associated with less absenteeism and thus higher effort than civil service teacher, better student performance, and given that contract teachers cost a fraction of what regular teacher cost the public exchequer, is a more cost-effective option (see Atherton and Kingdon (2010), Goyal and Pandey (2009), and Muralidharan and Sundararaman (2011)). A more recent meta-analysis of the issue of effectiveness of contract teachers compared to regular, civil service teachers by Kingdon et al (2013) suggests that contract teachers are indeed as effective, if not more effective, than regular civil service teachers in improving student learning outcomes though they caution against trying to generalize the results.

### **Block Grant Treatment**

Considerable attention has been paid to the importance of schooling inputs in ensuring educational outcomes. Given the overwhelming share of salary expenditures in government schools in India, the experiment aimed to determine returns to a student specific block grant<sup>12</sup> to schools to obtain teaching-learning materials for students. The program was run over two years. The per student block grant of

---

<sup>11</sup> Policy makers in India are not the only ones to advocate for the elimination of contract teachers. Well intentioned researchers globally have advocated for contract teacher elimination on the grounds that it is unethical to have different quality teachers for different students, that the salary differentials across civil service teachers and regular teachers leads to fragmented and demoralized staff, etc. See ILO 2012,

<sup>12</sup> Schools already receive block grants to support schools. They receive an annual grant of about INR 2000 for discretionary expenditures and about INR 500 per teacher for developing teaching learning material (TLM). However, given that there are typically about three teachers per rural school, we are comparing approximately INR 300,000 to each school for teacher salaries and INR 2500 plus the cost of supplying textbooks, uniforms, midday meals to each school as other inputs.

about INR 125 per year or given an average school size of about 80 students, translated to about INR 10,000 per school<sup>13</sup>. Grant money was typically used for procuring writing materials (40%), charts for classrooms (25%), workbooks and exercise books (20%) and about 10% on durables such as bags, plates and cups for the midday meal program, etc. The patterns of expenditure remain quite similar across years<sup>14</sup>.

One would therefore expect that learning outcomes trajectory to be similar to what is observed at the end of Year 1 but we see that this is not true for Year 2 results. It can also be seen from the nature of the expenditure, the items procured could have been provided by the parents rather than the institution.

### The Results

The study finds that students in schools that received a block grant had scores that were 0.09 and 0.08 SD higher than students in comparison schools in math and language respectively when comparing Year 1 results. These differences were significant. Students in schools that received block grants scored 0.04 SD and 0.07SD higher on math and language compared to students in comparison schools at the end of Year 2 and the difference is not significant for the end of Year 2 scores. Although the initial objective of the study was to merely study the impacts of school block grants aimed at teaching learning material for students on student achievement, we verified a result that mirrored earlier work by Das et al in Zambia. Household spending on education in program schools is significantly lower in Year 2 than in Year 1, suggesting that households were responding to the program by changing household spending patterns. That is, in Year 1, when the grant amount was unanticipated, the households were unable to or did not offset as much as they did when the grant amount had become anticipated by Year 2. The study shows that for each dollar of government spending in the form of a grant that the households now expect, there is a reduction of almost 85 cents in household spending on education.

### Policy Directions

The fact that households will re-optimize their budgets in response to public spending is something that policymakers and researchers have known for a long time. However, surprisingly, the vast literature on the impacts of schooling inputs on educational outcomes and on the estimation of education production functions seems to have overlooked such a response on the part of households. This suggests that policy makers may want to consider financing those inputs that households cannot easily substitute away from, for example, teaching inputs or school level infrastructure. At this point we are not advocating that governments stop financing those inputs that households can substitute away from by re-optimizing their household budgets.

### **Performance Pay**

The performance pay intervention was perhaps the most controversial and revolutionary treatment put into effect under this program. It was the first time in India when members of the bureaucracy (here

---

<sup>13</sup> The average spending for all four programs per school were calibrated to be equal.

<sup>14</sup> For a more detailed description of the provision of the block grant, please refer to Das et al (2012).

defined as government teachers) were being paid and rewarded for their performance<sup>15</sup>. Teacher pay for performance is not a new concept globally. While schooling inputs and resources available to schools have increased tremendously in recent years, this has not translated into improved student learning outcomes. Policy makers and researchers have therefore shown an interest in incentivizing teachers on the basis of a direct measure of the performance of their students. When APRESt was initiated, rigorous empirical evidence in the areas was limited. However, in the intervening 10 years this has become one of the most researched areas of education policy with numerous papers published in just the last three years<sup>16</sup>. Our results contribute to this growing body of literature and help to better understand how institutions and bureaucracies can be managed and made more efficient.

For any system that rewards on the basis of merit or performance several prerequisites are needed. These include *inter alia*: a clear set of goals, standards of performance on the basis of these goals that can be objectively and accurately measured, and finally, rewards provided on the basis of whether these standards were met. Alternatively, it is possible to construct a system where failure to meet established standards or norms are punished. The latter concept of punishing those who fail to meet certain basic norms does not exist in India and the evidence suggests that there is really no risk to an individual teacher for failure to perform or discharge their duties as a teacher adequately. Even basic measures of performance, such as teacher attendance, are not regularly and adequately monitored and measured, and rewarded or punished, let alone more outcome oriented aspects such as classroom performance and students learning outcomes. Teacher compensation is thus based solely on entry into service or seniority, unless the teacher breaks away from the core tasks of teaching and moves into administration and management.

Based on the comments of teachers during focus group interviews, subjective measures of performance often result in poor performers being rewarded. While teachers incentives seems like a logical way to align desired teacher behavior with the state's expectations, there are some concerns that need to be tackled with the design of incentive programs. These include: (i) incentives aligned to improved student test scores could get teachers "*teaching to the test*", rather than supporting a more wholesome development of the student; (ii) a second concern focuses on outright cheating on tests to improve student performance and hence teacher incentive payments, and (iii) perverse behavior on the part of teachers who can ensure that the strongest students in class are kept out of the baseline assessment and the weakest are kept out of class during the end line assessment and before incentive payments are calculated. The manner in which these concerns were addressed is described below.

*Teaching to the Test*: The program looked at this issue in two ways. Firstly, given that learning levels in rural government schools were very low, it was believed that teaching to the test may actually be an improvement in classrooms. Secondly, emphasis was placed on various aspects of test development. For example, the assessments were designed to measure performance on mechanical and conceptual

---

<sup>15</sup> There is a very cumbersome procedure for monitoring and measuring the performance of individual members of the bureaucracy. For the senior most officers, typically those from the premier civil service, the Performance Appraisal Report, is largely based on a results framework drawn up at the end of the year, with a rather subjective assessment by his or her immediate supervisor and reviewed by an even more senior member of the bureaucracy. There is no objective measure or criteria for performance.

<sup>16</sup> Glewwe, Ilias, and Kremer (2010), Martin (2010), Neal (2011), Behrman et al (2012), Contreras (2012), Fryer et al (2012), Springer et al (2012), Fryer (2013), and Goodman and Turner (2013).

questions<sup>17</sup>. Improvements on both types of questions allows us to conclude whether improved scores are purely due to teaching-to-the-test strategies adopted by teachers or are due to overall improved transactions in the classroom. While the program incentivized performances in Math and Telugu, during program implementation, student performance was also assessed on Science and Social Studies. Again, performance improvements on non-incentivized subjects also suggests broad-based learning under the program.

*Cheating in the Test:* With program like the No Child Left Behind and Race to the Top enacted in the USA, policymakers, teachers, administrators and others have expressed concern that the incentive structures and the high stakes testing results in perverse behavior on the part of teachers. In some cases, this perverse behavior results in outright cheating on the part of the teachers (Jacob and Levitt (2003))<sup>18</sup>. To ensure test score reliability, the assessments were carried out directly by one of the partners to the APRESt project, the APF, the main implementing partner. The APF was chosen for several reasons, most importantly the fact that they have a brand identity that is associated with promoting excellence in education and is an NGO respected by the teaching community. Furthermore, since it was not a direct stakeholder between improved results and pay outs for teachers, they were found to be the honest brokers that teachers were asking for early on in the study.

*Perverse Incentives Created:* to minimize incentives for perverse behavior, we tied the incentives to average improvements in child learning at the class/school level. So, if students were to drop out of the program after taking the baseline, that student receives a very low score and therefore teachers have a built in system to ensure minimal dropouts. Strengthening management of information systems, a key aim of the government, would also help ensure that teacher rewards were based on improvements demonstrated by a majority of the children and not by engineering wide margins by teachers.

Concern	Mitigation
Teaching to the test	Test designed in such a way that one could not do well without deeper knowledge / understanding. (Plus, given extremely low levels of learning, even test-taking itself is an important skill.)
Threshold effects <sup>19</sup>	Minimized by making bonus a function of average improvement of all students.
Neglecting weaker children	Incentives tied to <i>changes</i> from the baseline performance. Drop-outs assigned low scores, not ignored.
Cheating, paper leaks etc.	Assessment and grading of tests done by an independent 3 <sup>rd</sup> party.

### Incentive Design

We employed two types of incentives in the design. These were (i) monetary incentives paid to groups of teachers in schools (GI) and (ii) incentives paid on an individual basis to teachers in schools (II). The details of the incentive design are provided in the formal paper, however, here will simply state the nature of the incentive scheme. Teachers in GI and II schools were offered bonus payments on the basis of the

<sup>17</sup> Mechanical questions refers to those types of questions that either children are familiar with or directly test a concept. Conceptual questions refer to those set of questions that either are unfamiliar in how they are presented or more indirectly test the child's comprehension of a concept.

<sup>18</sup> Recent cheating on standardized assessments in Atlanta further underline concerns with high stakes testing and incentive programs based on the results of such assessments. However, these issues are not insurmountable and a reading of the Atlanta case suggests that even rudimentary checks and balances were not in place.

<sup>19</sup> Focusing only on students near an expected target or cut-off, and neglecting children far below and far above the cut-off or threshold.

average improvements in the Math and Language scores of students taught by them subject to clearing a minimum threshold of 5 percent<sup>20</sup>. All teachers in GI schools that received a bonus shared the bonus amount. In II schools, your class performance determined your bonus payment.

### The Results

From implementing the teacher performance pay experiment we find the following:

- (i) The study unambiguously demonstrates that GI and II led to improvements in student outcomes both over the initial two years of the program, and over the entire five years of the program. At the end of the first two years, students in incentive schools performed significantly better than their counterparts in comparison schools by 0.27 SD and 0.17 SD in Math and Telugu respectively. Over the entire five year span of the program, the magnitude of the effect was much larger for children in incentive schools with these students scoring 0.54 SD and 0.34 SD higher in Math and Telugu respectively than their counterparts in comparison schools.
- (ii) We do not find evidence that the incentive scheme resulted in any of the concerns identified earlier. We carry out robustness checks for teaching to test by decomposing the treatment effects by Repeat and Non-Repeat Questions, by Multiple-Choice and Free-Response Questions, and by Mechanical versus Conceptual Questions. In all cases, we find that children in incentive schools perform better than children in comparison schools suggesting broad based gains in incentive schools. Finally, we find that students in incentive schools also perform significantly better than children in comparison schools on non-incentive subjects – Science and Social Studies. At the end of two years, incentive school students score 0.11 SD and 0.18 SD higher than students in comparison schools on Science and Social Studies, while over five years this gap widens to 0.52 SD and 0.30 SD respectively.
- (iii) *Ex ante* we were unable to predict which one of these would perform better. Across five years of the program, students in II schools perform better than their counterparts in GI schools at every point over the five years, and while children in GI schools outperform their counterparts in comparison schools, though not significantly in every year. More importantly, there is no significant difference between GI and comparison schools over the five years. This result has now been replicated in several other studies and may begin to suggest that individual incentives might need to be reviewed carefully when thinking about teacher compensation policies.
- (iv) APRESt was initiated in an effort to boost teacher motivation and minimize teacher absence. At the end of two and five years of the program, we find no difference in attendance (or conversely absence rates) of students and teachers across incentive and comparison schools. Furthermore, we also do not find any significant difference across measures of classroom performance, such as, blackboard usage, encouraging classroom participation, etc. between incentive schools and comparison schools. We note that the superior performance in incentive schools is explained in part by enhanced teaching activity conditional upon presence in schools. In another study on the use of incentives, Duflo et al (2010) study the impact of incentives on attendance of teachers. They find that a simple piece rate scheme for attendance decreases teacher absenteeism in treatment schools by 21 percentage points

---

<sup>20</sup> Accounting for summer effects. This minimum 5 % threshold was removed in Year 2.

compared to teachers in comparison schools, and increases student performance by 0.17 SD. Therefore, clearly any restructuring of the compensation package for teachers might need to involve incentive on two different margins - one based on attendance and the other on learning outcomes. Furthermore, when teachers across incentive and comparison schools were asked *unprompted* questions on what they did differently after the end of the school year and before they knew whether or not they had received incentive payments for that year, teachers in incentive schools alluded to providing more practice examples, assigning more homework, stayed on in class beyond school hours to provide special assistance to the weaker students. These seem to be credible claims when these self-reported measures are correlated against the performance of their students, particularly given that less than half the teachers in incentive schools provide these reasons.

- (v) A direct comparison of the incentive programs and the treatments using unconditional inputs finds that the II schools spent about the same money as the schools with the input programs, though student scores in the II schools were three times the scores in schools with unconditional inputs. Though the effect sizes in GI schools were smaller relative to IIs, the program was still equally cost effective given the smaller bonus pay outs. The formal papers also look at two other ways to look at this issue of cost-effectiveness.

### Policy Directions

Several issues emerge from this work on performance pay. First, teachers are looking for ways by which an objective measure of their performance can be assessed and they can be rewarded accordingly. Secondly, and perhaps more importantly, the good and effective teachers know who they are, though the system as it exists at present is not able to recognize well performing teachers. In a separate paper on how teachers responded to the program, we find that the extent of teachers' stated ex-ante support for the program is positively correlated with their ex-post performance as measured by estimates of value addition. This may suggest that well designed performance pay systems will allow for teachers to sort themselves and for high performers to be recognized and rewarded and retained in the system. Even if performance is not rewarded in the manner in which it has been highlighted in this particular experiment, there is no question that measures of performance management need to be introduced into the system. This can be done in several ways. Perhaps the most important thing for the government to do would be to introduce student assessments on a regular basis and make learning outcomes a key system indicator which at present it is not.

## Section 5: Summary and Conclusions

APRESt represents a serious political and technical effort on the part of the GOAP and its partners to understand and analyze factors that could contribute to improved learning outcomes in classrooms in government schools across the state. Through a sustained period of approximately ten years, the program helped push both policy frontiers and academic boundaries in an effort to fully understand what works and what does not work in the improving learning outcomes in rural primary schools. The program pushed policy boundaries by its choice of methodology in which a state government in India allowed for a program to be randomized and thereby allowing rigorous, causal findings to be obtained. The program also pushed policy boundaries by experimenting with a set of policy options that are typically refuted even by well-meaning stakeholders purely on ideological grounds – such as performance pay and vouchers to support school choice. Finally, the program pushed policy boundaries by bringing in administrative, technical and academic teams around one table to try and address a question of global importance. While the findings from this study may not necessarily translate into other environments in India or even across the entire state of Andhra Pradesh, by experimenting on a large scale with rural government run institutions and by using mechanisms that closely mirror existing administrative set ups the program has set the stage for a larger scale roll out or at the very least a more intensive pilot. The program has also seriously pushed academic boundaries. The working group adopted an approach that aimed to ensure that this program would be of the highest academic quality. The best measure of academic quality is publishing the findings in peer reviewed journals. Till date four papers have already been published from APRESt with several other publications in the pipeline.

There are several takeaways from the research findings:

Evidence Based Policies: While much has been written about evidence based policies, rarely do you find governments adopting policies after a careful review of the evidence. Large scale prospective randomized evaluations allow governments to develop test beds for further policy research and development and then closely link the results of these evaluations to program implementation. As the education sector in India grows and matures, and the needs deepen, the government will need to find creative ways to achieve desired results and finance such programs. Development of cost-effective policies and strategies would be essential and the room for ideology driven policy development will need to shrink. Andhra Pradesh has taken the first step in this direction through the implementation of APRESt.

Enhanced Monitoring: As the paper on the impacts of diagnostic feedback to teachers demonstrates, monitoring alone is unlikely to result in improved classroom transactions and to learning outcomes across schools. Enhanced monitoring will have to be combined with punitive measures or rewards to ensure that desired outcomes are met. Given the political clout of teachers unions, and the understandable reluctance of democratically elected leaders to tangle with these unions, combining monitoring of learning outcomes and teacher effort, under a framework of reward and professional standards might be the best way to achieve improved learning outcomes in government schools in the country.

Focusing on Early Grades: There is a need to emphasize learning in early grades. In particular, children should demonstrate age appropriate reading and numeracy skills. Reading skills in particular are likely to

open up avenues for self-learning in later years. This then implies that norms such as pupil teacher ratios or children per classroom might need to be revised to allocate greater resources in early grades.

Learning Assessments: There is a need to strengthen assessment systems of student learning. While student assessments should definitely not be the sole basis for measuring the performance of an educational system, it should play an important part. If children are simply pushed through the system with no emphasis on learning, this could have disastrous consequences both for the individual (when trying to join a discriminating labor market) and for overall economic growth. Since India has moved to an era of low-stakes tests, and automatic promotion, there is a need to strengthen the monitoring and measurement of learning outcomes to ensure that learning levels are rising from the current low levels.

The Incentives Worked: Incentives work. While external factors are known to crowd out intrinsic motivation, we believe this is question of both how issues are framed and on how the incentives are designed. In the current setting, intrinsic motivation is crowded by a system that fails to recognize and appropriately reward effort and success. By bringing together a reward mechanism that is truly merit based, objective and understandable, it is possible to ensure that external rewards not only do not crowd out intrinsic motivation but reinforce them. We are not sure at this point whether the program would demonstrate the same results if run as a tournament as opposed to a contractual system established under APRESt. However, irrespective of the nature of the incentives framework in which a teacher is placed, it is important to monitor and measure performance of frontline providers. This is the only way by which the system can ensure that the best performers are attracted, invested in and retained in the system, while poor performers will self-select an exit strategy. Unfortunately, the current systems paints both higher performers and low performers with the same brush at best and at worst, actually rewards low performers. This inability to distinguish between them is highly demotivating.

Contract Teachers: This is perhaps the most controversial finding of the research program. Well intentioned and meaning stakeholders to the education system often find themselves on opposite ends of the spectrum when it comes to the definition of teacher quality and how one defines teacher quality. Rigorous research findings now clearly demonstrate that observable teacher characteristics have little to do with how well or how poorly their students perform. That is, licensed teachers do not necessarily do better than unlicensed ones, teacher with higher qualifications do not necessarily do better than teachers with a lower level of educational attainment, performance of teachers tends to plateau out in a relatively short period of time, and in the context of APRESt, poorly qualified and trained contract teachers from this study (and others) seem to do at least as well a highly qualified, tenured civil service teacher in raising student learning outcomes. However, the RTE as enacted calls for a ban on contract or para teachers and that these should be replaced with full time, civil service tenured teachers. The results of this and other studies might suggest a review of this part of the RTE particularly given the enormous fiscal implications of this provision.

## References

Please see attached papers.