To the NEUDC 2006 Paper Selection Committee,

The paper I am submitting for NEUDC 2006 is based on a multi-year project in India, from which the data is still coming in. As a result, I have done a fair bit of analysis, but will only finish writing the paper in July.

While it is possible to submit an extended abstract, I thought that I can convey a much better flavor of the full paper with a presentation than in a 1000-word extended abstract . I have discussed this with Prof. Kaushik Basu (chair of the program committee) and he told me that this would be fine and that I should include a note to this effect in my submission.

I trust that this will be an effective way for the conference committee to get a sense of what the full paper will look like. A full draft will be available to send to the discussants by the end of July.

Regards,

-Karthik

Teacher Incentives in Developing Countries: Experimental Evidence from India

Karthik Muralidharan
Department of Economics, Harvard University

(with Venkatesh Sundararaman, World Bank)

Outline of Presentation for NEUDC 2006

Agenda

- **Overview**
- Experimental Design
- Results
- Conclusions and Next Steps

Main Results

- Teacher incentives (bonus payments based on average improvement in test scores) had a strongly significant positive impact on student learning outcomes measured by test scores
  - -Math and Language scores up by 0.15 and 0.09 standard deviations in incentive schools relative to control schools
  - -Significant positive effect in all grades (1-5)

- No significant difference between group and individual incentive programs
  - -But the magnitude of effects is larger with individual incentives, and the mean school size is small

- No adverse effects were detected in Incentive Schools
  - -Improvement in "conceptual" as well as "mechanical" questions in both Math and Language tests
  - -Better classroom process indicators in incentive schools

- This is likely to be a lower bound on the effect of incentives
  - -The measured difference is based on the "announcement" of the incentive program and no payments have been made yet
  - -Effect likely to be larger in year 2 of the program after payments in year 1

Background

- Large inefficiencies in service delivery in developing countries (especially in health and education)
  - -In India, 25% teachers are absent, less than half are teaching

- Following discussions of the findings from the teacher absence work with government officials
  - -We offered to design and evaluate the effectiveness of a teacher incentive program in government primary schools in the Indian state of Andhra Pradesh

- This is part of a multi-year randomized evaluation of various educational interventions
  - -First year of program (2005-06) featured 2 input treatments and 2 incentive treatments
  - -This paper focuses on the impact of the 2 teacher incentive programs (group and individual incentives)

Literature

- Teacher incentives
  - -Lavy (2002) and (2004) in Israel
  - -Glewwe, Ilias, Kremer (2003) in Kenya
  - -Duflo and Hanna (2005) in India
  - -Jacob (2005) and others in the US

- Incentives in public service delivery in developing countries
  - -World Development Report (2004)
  - -Das and Hammer (2005)
  - -Banerjee, Deaton, and Duflo (2005)

- Multi-task moral hazard/Incentives in Organizations
  - -Holmstrom & Milgrom (1991)
  - -Baker (1992) and (2002)

- Effectiveness of Aid in Development
  - -Sachs (2005), Easterly (2006)

Main Contributions

- First randomized evaluation of teacher incentives in a representative sample of government/public schools in either a developing or developed country

- High internal and external validity of results
  - -Internal validity via random assignment of treatment
  - -External validity via stratified random sampling of study universe

- Take test-design seriously and include both mechanical and conceptual test questions to get at concerns regarding teaching to the test

- Study group and individual teacher incentives in the same field experiment

- Variation in base pay of teachers allows us to estimate the impact of the magnitude of the financial incentive

Agenda

- Overview
- **Experimental Design**
- Results
- Conclusions and Next Steps

Project Timeline

- Phase 0: Jan – Dec 04 (discussions with the Government of Andhra Pradesh, proposal refinement, extensive incorporation of expert feedback, funding, partner selection, MoU signing)

- Phase 1: Dec 04 – March 05 (Pre-testing and calibration, refining of processes – process pilot in 40 schools in one district)

- Phase 2: June 05 – May 06 (Main Study across 5 districts)

  -Reporting Results from Phase 2 in this paper

- Phase 3: June 06 – May 07 (Sustainability and validation of findings after credibility of program is established)

- Phase 4: June 07 onwards (Tracking of longer term outcomes and various project extensions)

Incentive Design

- Monetary incentive to be paid on the basis of average improvement of test scores of all students in grade/school over base line
  - -Improvements rather than absolute levels because they provide a rationale to improve performance of ALL students, without any threshold effects

- Based on absolute improvement (contract/piece rate) as opposed to improvement relative to other schools (tournament)
  - -Common shocks unlikely as schools are spread out across state
  - -Contracts dominate for risk-averse agents in this setting
  - -The main reason for preferring contracts was because they are more transparent when the treatment universe is not known to the schools
  - -Advantage of tournament is budgetary predictability
    - Might be a better model if the idea is scaled up and all schools participate

- Linear incentive schedule in improvement with a minimum threshold of improvement before the bonus kicks in
  - -The threshold was used because children would perform better at the end of the year simply because they are older (also June to March testing effects)

Incentive Details

- Calibrations from the pilot showed that the "business as usual" improvement between grades 3 and 4 on the SAME grade 3 test was ~6-7% points

- We therefore set the improvement threshold at which the incentive payment "kicks in" to be 5% points, followed by a Rs. 500 bonus payment for every 1% point improvement in average scores.
  - -Individual incentives: Each teacher to receive a bonus based on the average improvement shown by children taught by that teacher
  - -Group incentives: All teachers to receive an equal bonus based on the average improvement of ALL the children in the school

- Thus a 10% point improvement in scores (0.4 standard deviations of pilot score distribution) would lead to a bonus payment of Rs. 2,500 (one third of one month's salary for a regular teacher)
  - -A 5% improvement would lead to no bonus payment
  - -Ceiling effects were not a problem

Testing Details

- Baseline tests based on previous year's competencies
  - -Each subject test has 25-35 questions mapped into 5-7 competencies

- Endline tests the same competences (in Math & Language) to enable accurate comparison and measurement of improvement
  - -Also included additional content from the current year's competencies

- To preserve similarity of testing conditions between baseline and endline, we break the endline testing into 2 days:
  - -Endline: Covers baseline competencies (up to previous years competencies)
  - -Higher Endline: Current year competencies (and non-incentive subjects)

- The test is designed to be able to distinguish between "no learning", "mechanical learning" and "conceptual learning" for each student
  - -More on this later

Summary of Experimental Design

- Study conducted across 500 primary schools in 5 districts of AP
  - -5 districts to represent all 3 distinct socio-cultural regions of AP
    - 10 randomly selected sub-districts in each district (in same division)
      - 10 randomly selected schools in each sub-district (mandal)

- Conduct independent testing in these 500 schools (June/July 05)

- Provide diagnostic feedback on test performance to all schools and announce different programs in different schools (August 05)
  - -100 schools **randomly assigned** to each of 4 treatments, 100 in control
  - -Stratified random allocation (2 schools in each mandal to each "treatment")
  - -Randomization and announcement of treatments were done AFTER the baseline to eliminate possibility/effects of gaming baseline scores

- Carefully monitor various aspects of school-performance over the course of the year via unannounced monthly tracking surveys
  - -Collect data from all stake holders (parents, children, teachers)
  - ~50,000 children tracked over this period (including a household survey)

- Conduct independent end of year tests to assess the impact of various interventions on learning outcomes (March/April 06)

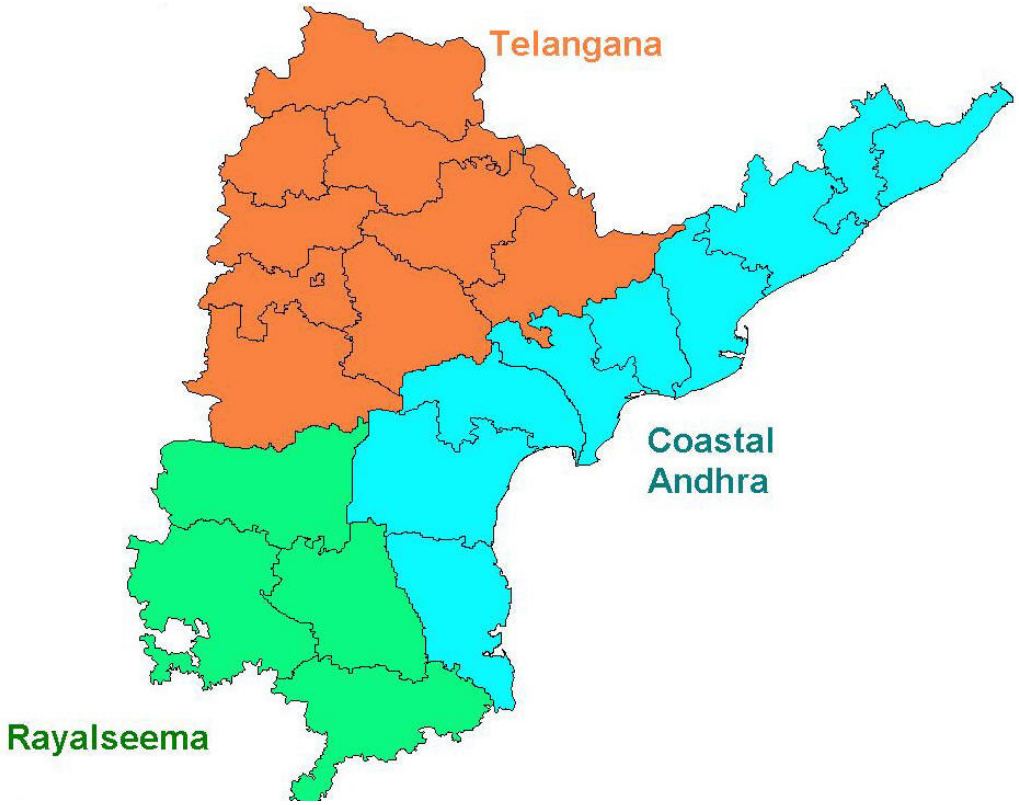
Location of Study

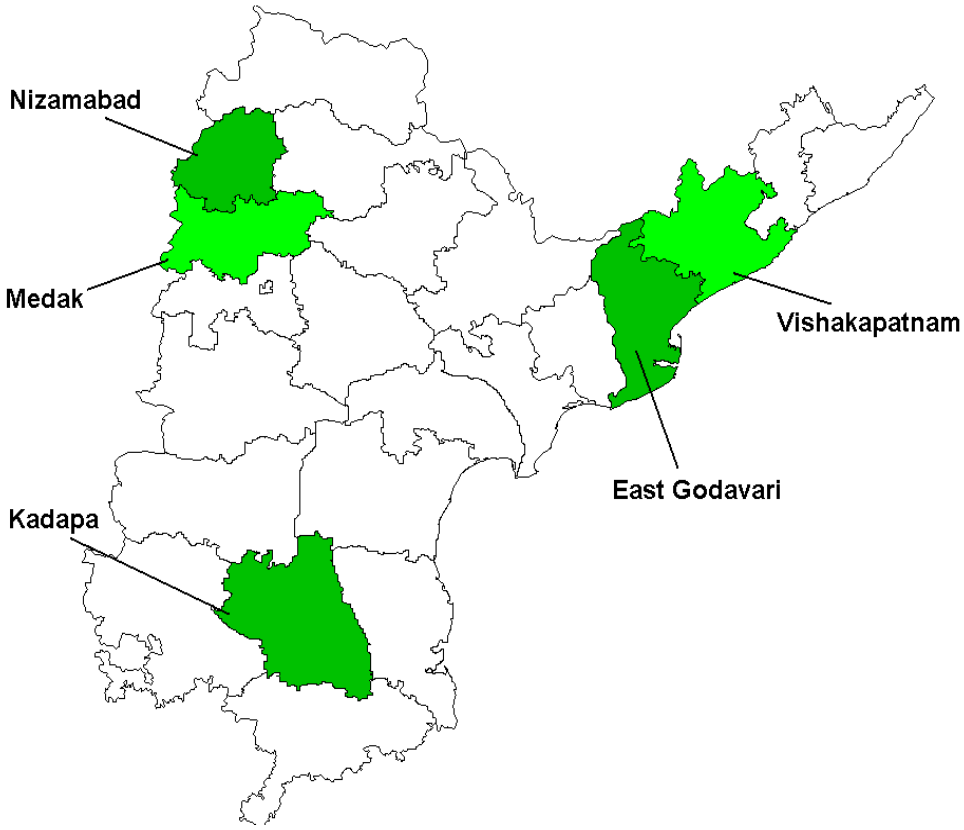
- Indian State of Andhra Pradesh (AP)
  - -5th most populous state of India
    - Population of 75 Million
  - -23 Districts (2-4 Million each)

- Close to All-India averages on many measures of human development

	India	AP
Gross Enrollment (6-11) (%)	95.9	95.3
Literacy (%)	64.8	60.5
Teacher Absence (%)	25.2	25.3
Infant Mortality (per 1000)	63	62

District Sampling and External Validity

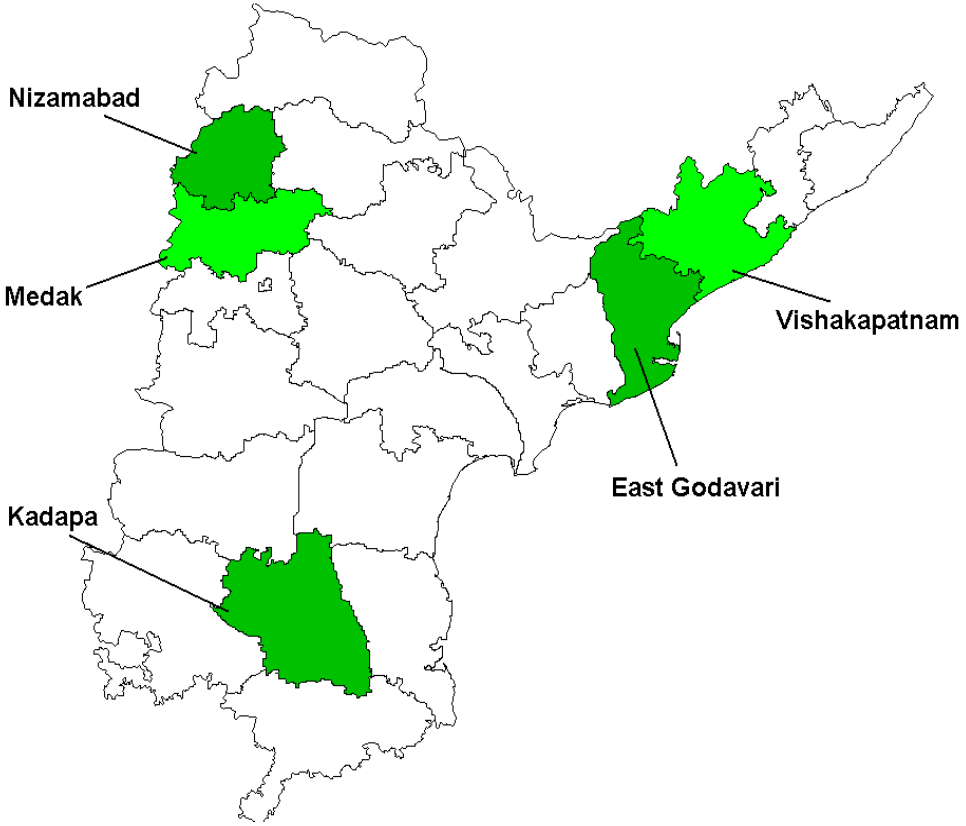




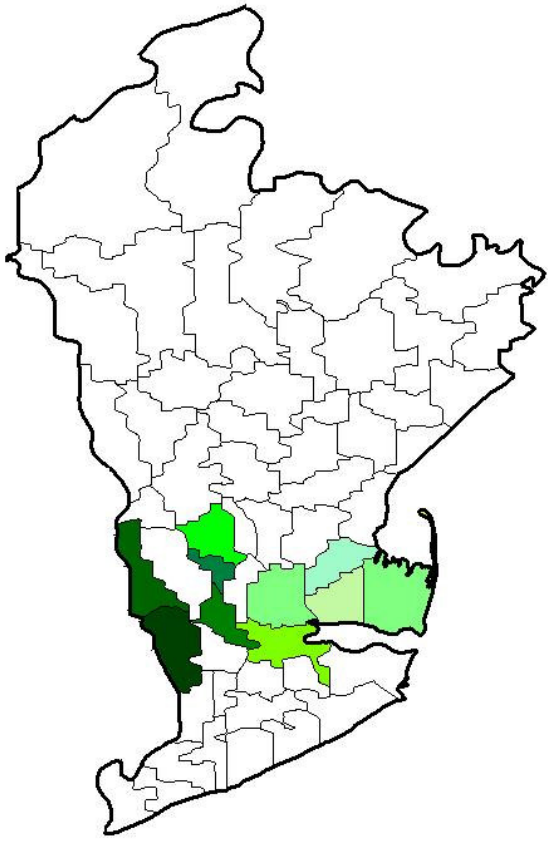
- Districts chosen from each of the three distinct socio-cultural regions of the state proportional to population (chosen contiguously where 2 districts were chosen)

- Increases external validity of the study since every government-run rural primary school in the state had a similar ex ante probability of being chosen for the study

Sub District (Mandal) Sampling Example



Andhra Pradesh



East Godavari District

- One "division" chosen randomly in each district and ten "mandals" chosen randomly in the sampled division.
- The mandal is the lowest administrative unit in the state (including education)

Final Design Overview

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

- Stratified random assignment of treatment –
  -2 schools in each mandal assigned to each treatment arm.
- Thus each mandal is a microcosm of the entire study (across 50 mandals in 5 districts) and statistical power is increased by including mandal fixed-effects

Agenda

- Overview
- Experimental Design
- **Results**
- Conclusions and Next Steps

Sample Balance

Sample Perfectly Balanced on Observables Across Treatments

	Control	Extra Para Teacher	Block Grant	Group Incentives	Individual Incentives	Largest Difference	P-Value (for largest difference)
School Enrolment	102.0	99.9	95.4	99.9	105.0	9.6	0.25
Pupil Teacher Ratio	28.4	30.5	26.6	28.3	29.0	3.8	0.10
Average Number of Teachers	3.1	2.8	3.0	3.1	3.1	0.3	0.14
Baseline Telugu Score	35.4	34.9	33.8	34.7	33.9	1.6	0.31
Baseline Maths Score	18.4	17.3	17.0	17.4	17.7	1.4	0.22

Specification

$$y_{ijkm} = \alpha + \beta \cdot Z_m + \gamma \cdot Incentives + \varepsilon_i + \varepsilon_j + \varepsilon_k$$

$$i = Child\,,\ j = Class\,, k = School\,, m = Mandal\,(Sub - District\,)$$

y = Normalized test score gain (by the std. dev. of endline test score distribution in each specific test – e.g. Grade 4 Math)

- All regressions include Mandal fixed effects, and standard errors are clustered at the school level

Impact of Incentives on Test Scores

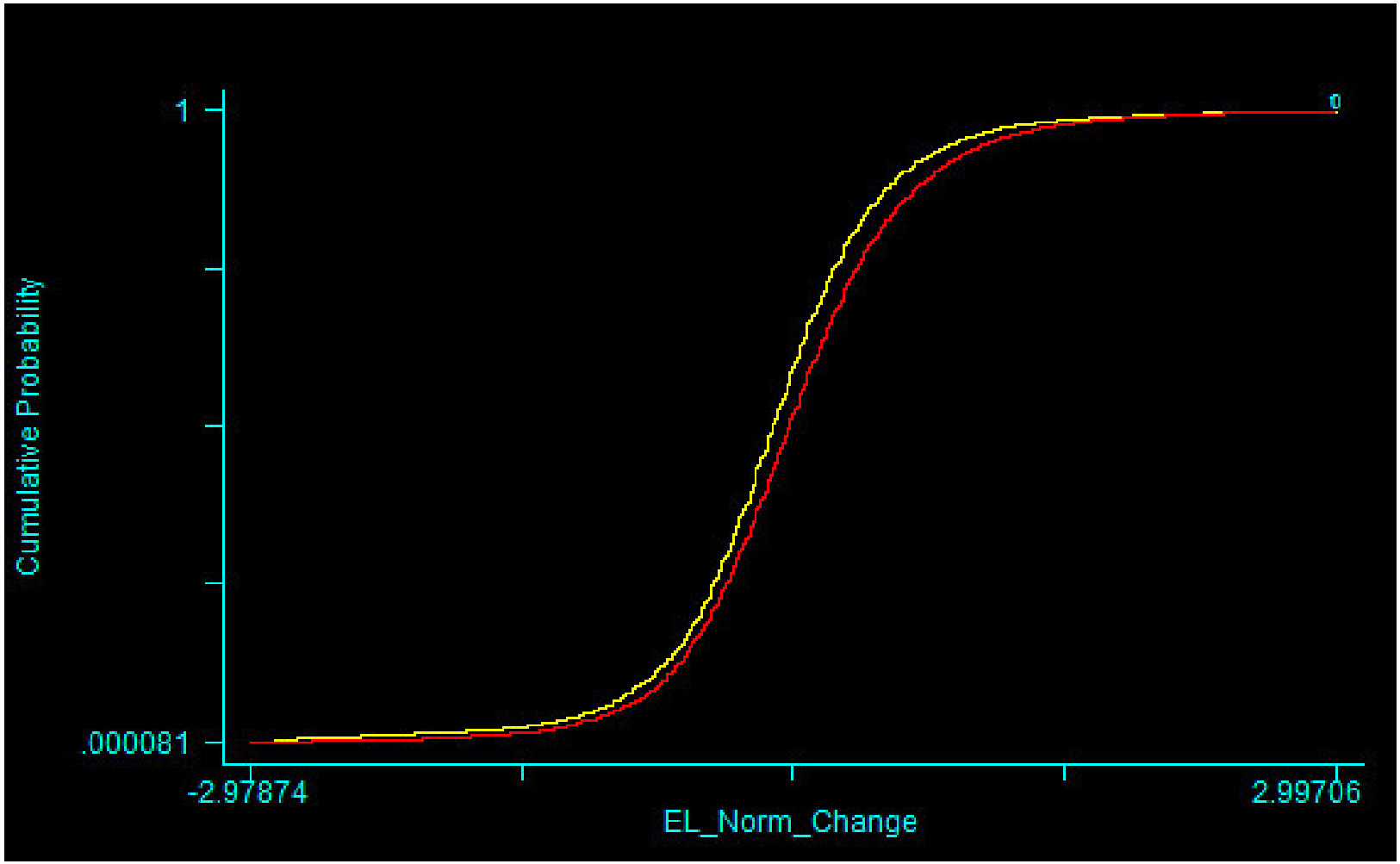
Table 1: Impact of Incentives on Learning Outcomes						
	Dependent Variable = Normalized Change in Test Scores by Child					
	Combined		**Math**		**Telugu (Language)**	
	[1]	[2]	[3]	[4]	[5]	[6]
Incentives (Pooled)	0.115 (0.028)***		0.145 (0.038)***		0.085 (0.025)***	
Group Incentive		0.102 (0.034)***		0.135 (0.045)***		0.069 (0.029)**
Individual Incentive		0.127 (0.033)***		0.154 (0.044)***		0.1 (0.028)***
Constant	-0.086 (0.023)***	-0.086 (0.023)***	-0.108 (0.031)***	-0.108 (0.031)***	-0.063 (0.021)***	-0.063 (0.021)***
Observations	37117	37117	18435	18435	18682	18682
R-squared	0.06	0.06	0.07	0.07	0.07	0.07
All Regressions include mandal (sub-district) level fixed effects Robust standard errors clustered at the school level in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%						

Impact of Incentives by Grade

Table 2: Impact of Incentives by Grade						
	Combined		Math		Telugu (Language)	
	[1]	[2]	[3]	[4]	[5]	[6]
Incentives * Grade 1	0.12	-0.003	0.151	-0.005	0.09	-0.001
	(0.059)**	-0.063	(0.080)*	-0.086	(0.048)*	-0.053
Incentives * Grade 2	0.079	-0.045	0.093	-0.063	0.064	-0.026
	(0.034)**	-0.038	(0.043)**	-0.048	(0.034)*	-0.039
Incentives * Grade 3	0.134	0.011	0.166	0.01	0.103	0.012
	(0.037)***	-0.031	(0.047)***	-0.042	(0.034)***	-0.033
Incentives * Grade 4	0.109	-0.015	0.143	-0.013	0.074	-0.016
	(0.036)***	-0.031	(0.048)***	-0.04	(0.031)**	-0.032
Incentives * Grade 5	0.123		0.156		0.09	
	(0.036)***		(0.049)***		(0.032)***	
Incentive Treatment		0.123		0.156		0.09
		(0.036)***		(0.049)***		(0.032)***
Constant	-0.086	-0.086	-0.108	-0.108	-0.063	-0.063
	(0.023)***	(0.023)***	(0.031)***	(0.031)***	(0.021)***	(0.021)***
Observations	37117	37117	18435	18435	18682	18682
R-squared	0.06	0.06	0.07	0.07	0.07	0.07

All Regressions include mandal (sub-district) level fixed effects
Robust standard errors clustered at the school level in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Cumulative Distribution of Gains by Treatment



Concerns with Incentives based on Test Scores

- Teaching to the test
  - -Less of a concern given extremely low levels of learning
  - -Test-taking is an important skill in the Indian context
  - -Test design can get progressively more sophisticated so that you cannot do well on the test without deeper knowledge/understanding

- Cheating/Paper leaks etc.
  - -Potentially a big problem – here the testing is done by a completely independent outside testing team
  - -Stringent test-integrity measures taken

- Teachers neglecting/penalizing weaker children or other adverse behavioral consequences of incentives
  - -Mitigate/avoid the former by tying incentives to "changes" from the baseline performance and assigning low scores to drop outs
  - -Test for latter by collecting data on many process variables

Mechanical Versus Conceptual Learning

- An important innovation of this project
- In each "competence" identified by Education Initatives (the testing firm), they include conceptual questions that test concepts that are supposed to be known in an unfamiliar way
- Of course, a "conceptual" question that is taught repeatedly in class can become a "mechanical" one
- Our operational distinction therefore is that questions that adhere to the patterns in the text books are considered "mechanical", while those that take the same underlying idea and ask the question in an unfamiliar way are considered "conceptual"
- Examples: Area, multiplication, estimation

Impact of Incentives by Mechanical/Conceptual

Table 3: Mechanical Versus Conceptual Learning						
	Combined		**Math**		**Telugu (Language)**	
	Conceptual	Mechanical	Conceptual	Mechanical	Conceptual	Mechanical
	[1]	[2]	[3]	[4]	[5]	[6]
Incentives (Pooled)	0.116 (0.035)***	0.118 (0.028)***	0.144 (0.045)***	0.151 (0.039)***	0.088 (0.034)**	0.084 (0.024)***
Constant	-0.086 (0.028)***	-0.088 (0.022)***	-0.112 (0.036)***	-0.111 (0.031)***	-0.06 (0.028)**	-0.064 (0.020)***
Observations	37111	37114	18432	18433	18679	18681
R-squared	0.05	0.05	0.06	0.06	0.05	0.07
All Regressions include mandal (sub-district) level fixed effects Robust standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%						

- Incentive schools had significant improvements in both the mechanical and conceptual components of the tests in both math and language.
- Gains in math are greater than gains in language, which also suggest that we are not merely picking up 'teaching to the test' since Math is harder to cram

Test Integrity

- We have no reason to believe that the integrity of the tests was compromised in any way
  - -Question papers in sealed packets by school
  - -Testing teams of 5 evaluators (one team leader) supervised by a sub-district coordinator (who was on the project team for the full year)
  - -Identity verification of children taking the test
  - -Standard test procedures (distance between test takers, continuous proctoring, teachers away from testing area, etc.)
  - -The assessments/coding of the papers were done at sub-district headquarters in the evening of each testing day under the supervision of the sub-district project coordinator

- Have looked at the distribution of extreme positive outliers by treatment category
  - -No significant difference across treatment categories in children improving by over 4 standard deviations
  - -Will be doing more analysis on this looking at school-level correlations in scores

Attrition

- Need to protect against incentive schools inflating average scores by encouraging weak children to drop out
  - -But also need to not penalize teachers too much for drop outs beyond their controls
  - -So we tell them that we will assign the same baseline score in the endline for children who are in the baseline but not in the endline
    - Effectively calculated as "no improvement"
    - Since there is a 5% threshold of average improvement before the incentive kicks in, this does penalize the teachers to some extent, but not extremely so

- We have attendance data at the child-level from our tracking surveys and will use this to measure attrition

- For now, if we consider children who took the baseline but did not take the endline as drop outs, we find that:

- Incentive schools have lower attrition:
  - -20.6% versus 21.4%

Process Variables (Classroom Observation)

Process Variables: Based on School/ Classroom Observation			
	Control	**Incentives**	**p-value**
Teacher Absence (%)	20.3	21.0	0.53
Active Teaching (%)	**43.9**	**46.2**	**0.09***
Active Blackboard Usage (%)	43.7	45.5	0.19
Writing on Blackboard Clearly Visible from back of Class (%)	**72.3**	**75.3**	**0.01****
Teacher Calls Students by Name (%)	**68.8**	**71.8**	**0.02****
Teacher Addresses Questions to Students (%)	**52.4**	**54.7**	**0.10***
Teacher Reads from Textbook (%)	46.7	50.1	0.21
Teacher Assigned Homework (%)	32.8	34.6	0.17

Process Variables (Household Interview)

Process Variables: Based on Household Interview			
	Control	**Incentives**	**p-value**
Teacher Corrects Workbook (%)	33.7	36.2	0.13
Teacher Provides you with Individual/ Small Group Help (%)	61.9	64.1	0.17
Teacher encourages you for good work or improvement (%)	**64.5**	**67.2**	**0.10***
Teacher makes you listen to the radio in school (%)	**42.5**	**46.1**	**0.04****
Has the teacher ever beaten you in school? (%)	73.2	71.1	0.18

Agenda

- Overview
- Experimental Design
- Results
- **Conclusions and Next Steps**

Findings So Far

- Teacher incentives (bonus payments based on average improvement in test scores) had a positive impact on student learning outcomes measured by test scores
  - -Math and Language scores up by 0.15 and 0.09 standard deviations in incentive schools relative to control schools
    - A median child in an incentive school would place around the 55^{th} – 60^{th} percentile of a control school distribution (varying by grade/subject)
  - -Significant positive effect in all grades (1-5)

- No significant difference between group and individual incentive programs
  - -But this needs more digging into

- No adverse effects were detected in Incentive Schools
  - -Improvement in "conceptual" as well as "mechanical" questions in both Math and Language tests
  - -Math Δ > Language Δ also supports the idea that we are not just picking up "teaching to the test" since Math is harder to cram
  - -Better classroom process indicators in incentive schools

Next Steps

- Impact of incentives as a function of teacher base pay
  - Differential impact of incentives by teacher type and demographics
- Differential impact of group and individual incentives as a function of school size and ratio of regular teachers to para-teachers
- Robustness of results to outliers and careful examination for class/school level cheating
- Impact of incentives on performance on additional subject (EVS) that was tested but was not a part of the incentive plan
- Heterogeneity of treatment effects by demographic characteristics
- Do a cost-benefit analysis and compare with other interventions
- Data is still being entered and cleaned, and should be fully ready by middle of June. Full paper draft will be ready by end of July.