Pay by Design

Teacher Performance Pay Design and the Distribution of Student Achievement

Sean Sylvia (Renmin University of China)

with

Prashant Loyalka (Stanford University)

Chengfang Liu (Chinese Academy of Sciences)

James Chu (Stanford University)

Yaojiang Shi (Shaanxi Normal University)

We thank the Ford Foundation and the Xu Family Foundation for funding this study.

RENMIN UNIVERSITY OF CHINA

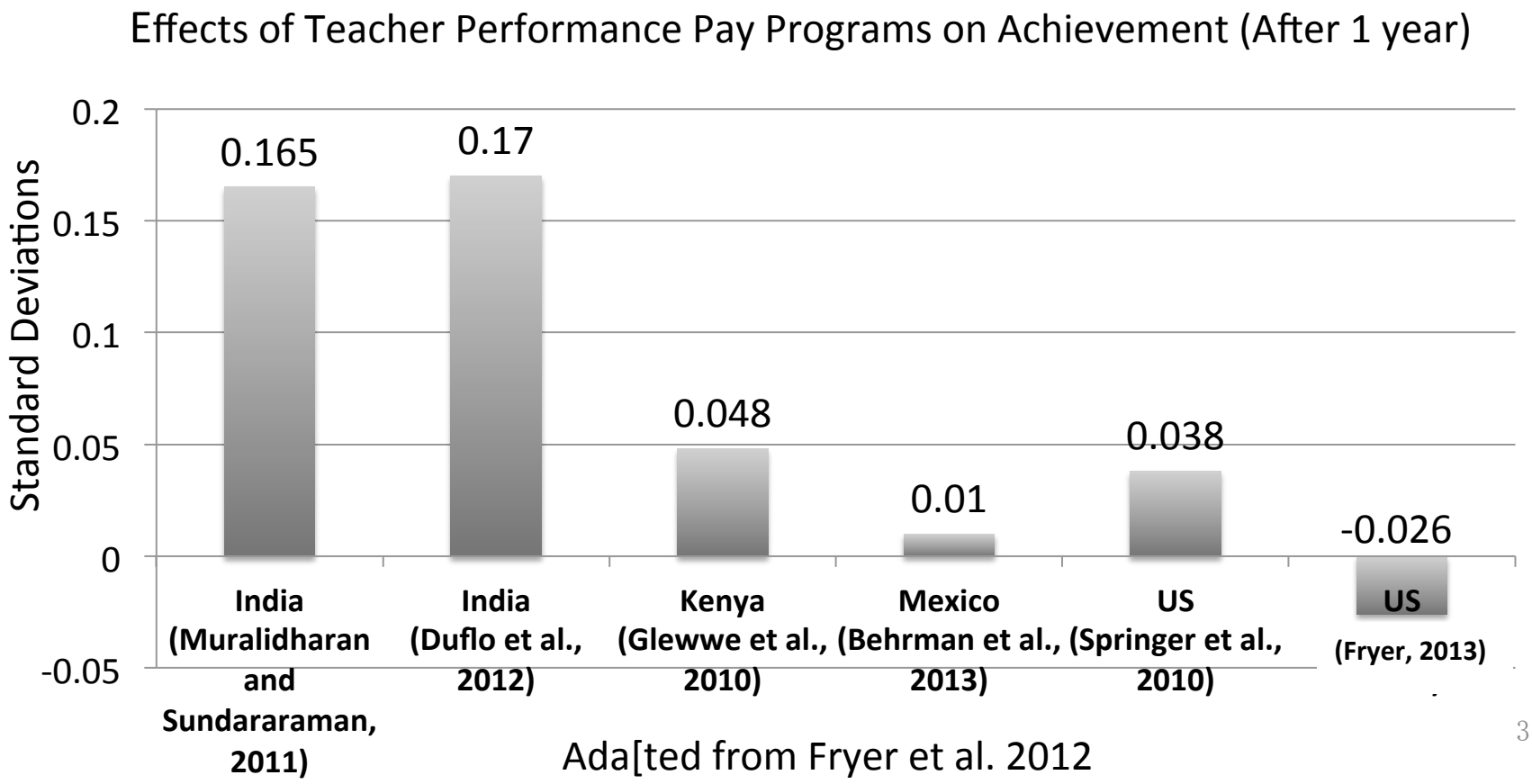
REAP Rural Education Action Program

Teacher Performance Pay

- Teachers among most important inputs to student achievement (Aaronson, Barrow, and Sander, 2003; Rockoff, 2004; Rivkin, Hanushek, and Kain, 2005; Hanushek and Rivkin, 2010; Rivkin, 2006; Chetty, Friedman and Rockoff , 2013)
- But often work in settings where they face incentives that are weak or misaligned with improving student outcomes (Lazear, 2003)

Teacher Performance Pay

- Widespread policy interest in motivating teachers by linking pay to performance metrics – commonly student exam scores:
  - US, Australia, UK, Israel; Mexico, Chile, Kenya, India, Pakistan, **China**
- But, mixed evidence on effectiveness



Performance Pay Design

- One reason for mixed evidence on teacher performance pay may be because the design of performance pay varies across studies (Neal, 2011)

- Two design features which vary across studies:
  - **Design feature 1:** the way in which student achievement scores are used to measure teacher performance & mapping to rewards
  - **Design feature 2:** the size of the rewards

- Despite the theoretical importance of these design features, there is little empirical evidence about how varying them affect:
  - Student achievement on average
  - The achievement of different types of students

- Theoretically compelling designs may not outperform simple/ transparent schemes in practice

This Study

Randomized trial across 216 primary schools in rural western China to study

1. How different ways of using student achievement to measure and reward teacher performance affect teacher effort and student achievement

2. Whether the size of potential rewards matters

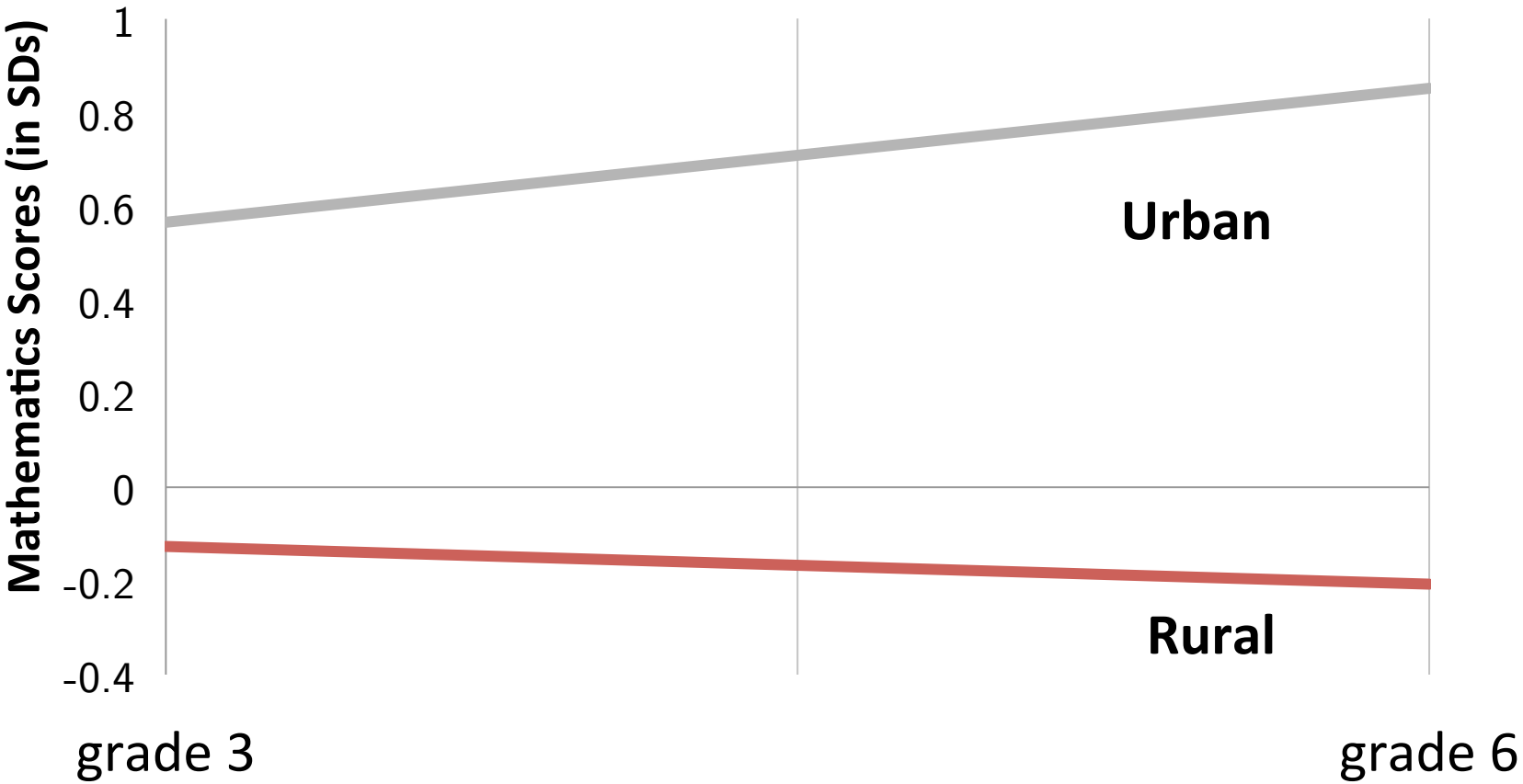
3. How different performance pay designs affect achievement among low, medium, and high achieving students within the classroom? (i.e. do teachers "triage" students in response to incentives)

Rest of the Presentation

- **Background/Context**
  - **Study in Rural China**
  - **Teacher Performance Pay Policy in 2009**
- Experimental Design and Interventions
- Sampling/Data
- Results + Discussion

Rural China: low levels of learning

- Rural-urban achievement gap grows as children progress through the education system (0.8 SD gap in Math by grade 6).

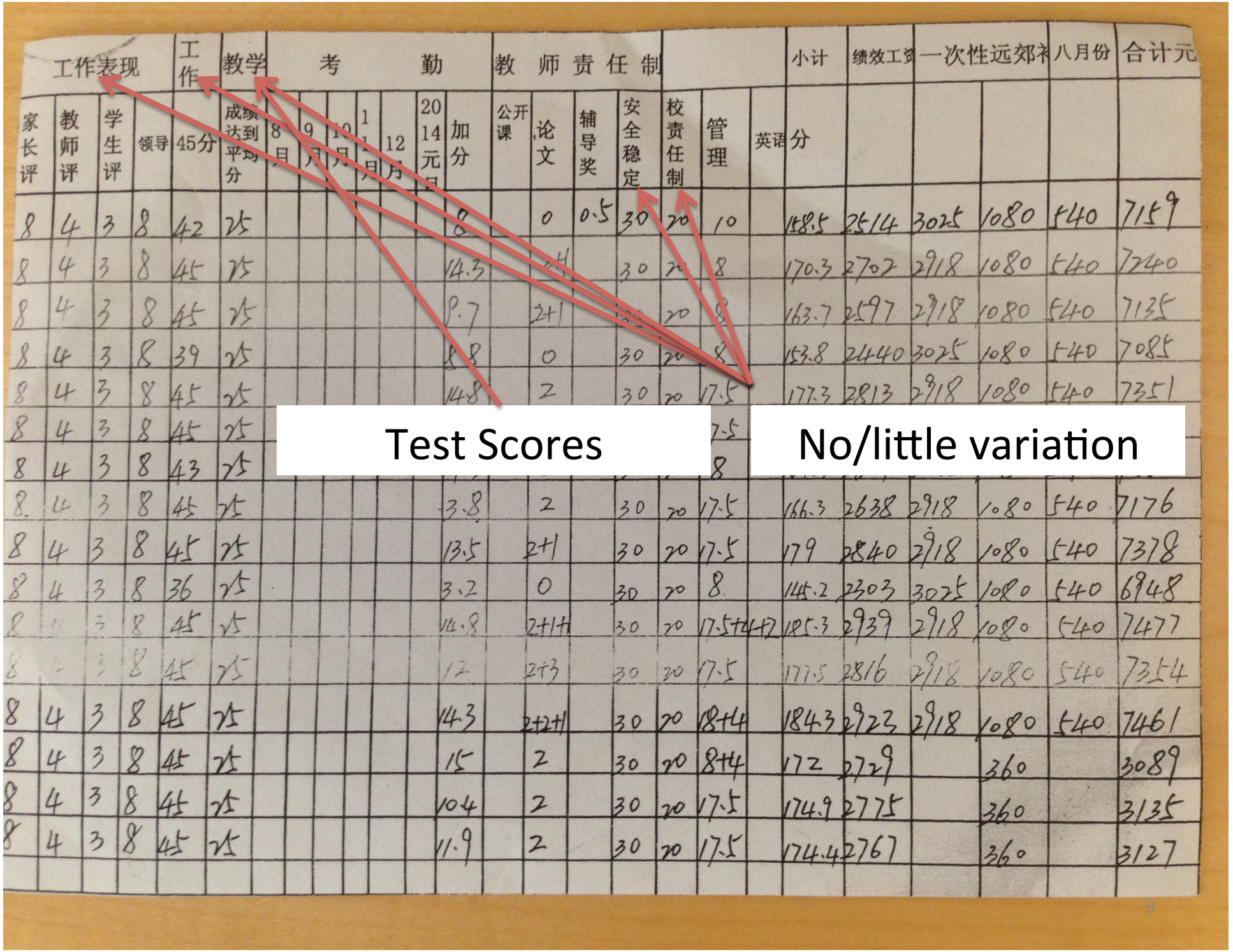


Low levels of learning despite large, large increases in government expenditures on rural, compulsory education (NBS, 2011)

Teacher Performance Pay Policy in China

- 2009 Teacher Performance Pay Policy
  - Increased teacher salaries to the level of other local civil servants
  - Stipulated that 30% of increase be awarded based on performance

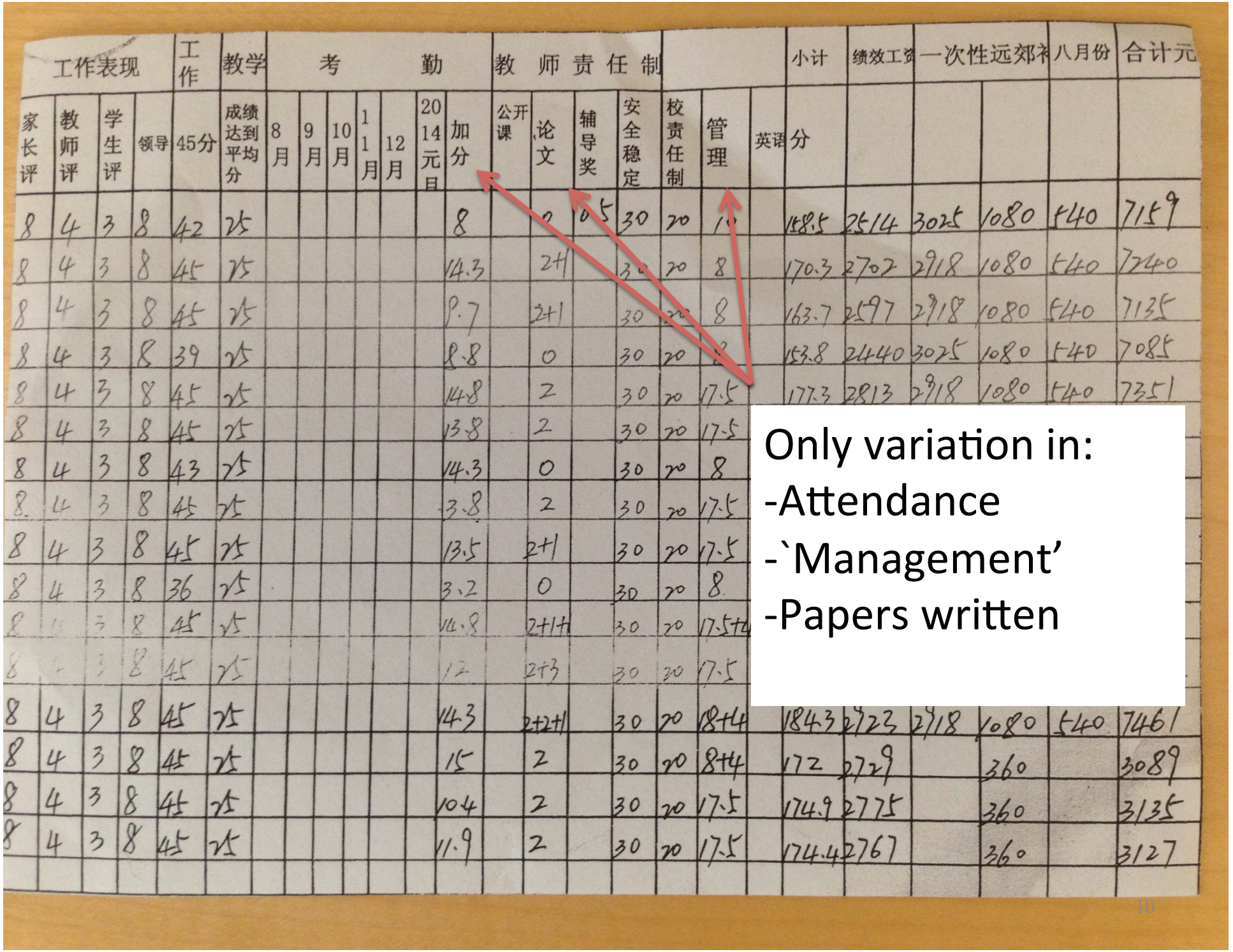
- **How was the policy actually implemented?**
  - Teacher performance based mainly on inputs (e.g. class hours) and subjective measures
  - **Little variation in actual rewards: 300 yuan difference per semester between top and bottom teacher on average**
  - **Teachers rankings done WITHIN schools (potentially problematic)**
  - When evaluated on student scores, rankings based on levels



工作表现				工作	教学	考勤						教师责任制							小计	绩效工资	一次性远郊补		八月份	合计元
家长评	教师评	学生评	领导	45分	成绩达到平均分	8日	9月	10月	11月	12月	2014元旦	加分	公开课	论文	辅导奖	安全稳定	校责任制	管理	英语	分					
8	4	3	8	42	25							0		0	0.5	30	20	10		158.5	2514	3025	1080	540	7159
8	4	3	8	45	25							14.3		4		30	20	8		170.3	2702	2918	1080	540	7240
8	4	3	8	45	25							8.7		2+1			20	8		163.7	2597	2918	1080	540	7135
8	4	3	8	39	25							8.8		0		30	20	8		153.8	2440	3025	1080	540	7085
8	4	3	8	45	25							14.8		2		30	20	17.5		177.3	2813	2918	1080	540	7351
8	4	3	8	45	25													17.5							
8	4	3	8	43	25													8							
8	4	3	8	45	25							13.8		2		30	20	17.5		166.3	2638	2918	1080	540	7176
8	4	3	8	45	25							13.5		2+1		30	20	17.5		179	2840	2918	1080	540	7378
8	4	3	8	36	25							3.2		0		30	20	8		145.2	2303	3025	1080	540	6948
8	4	3	8	45	25							14.8		2+1+1		30	20	17.5+4+2		185.3	2939	2918	1080	540	7477
8	4	3	8	45	25							12		2+3		30	20	17.5		177.5	2816	2918	1080	540	7354
8	4	3	8	45	25							14.3		2+2+1		30	20	18+4		184.3	2923	2918	1080	540	7461
8	4	3	8	45	25							15		2		30	20	8+4		172	2729		360		3089
8	4	3	8	45	25							10.4		2		30	20	17.5		174.9	2775		360		3135
8	4	3	8	45	25							11.9		2		30	20	17.5		174.4	2767		360		3127

Test Scores

No/little variation



工作表现				工作	教学	考勤							教师责任制							小计	绩效工资	一次性远郊补		八月份	合计元
家长评	教师评	学生评	领导	45分	成绩达到平均分	8月	9月	10月	11月	12月	2014元旦	加分	公开课	论文	辅导奖	安全稳定	校责任制	管理	英语	分					
8	4	3	8	42	25							8		0	0.5	30	20	15		168.5	2514	3025	1080	540	7159
8	4	3	8	45	25							14.3		2+1		30	20	8		170.3	2702	2918	1080	540	7240
8	4	3	8	45	25							8.7		2+1		30	20	8		163.7	2597	2918	1080	540	7135
8	4	3	8	39	25							8.8		0		30	20	8		153.8	2440	3025	1080	540	7085
8	4	3	8	45	25							14.8		2		30	20	17.5		177.3	2813	2918	1080	540	7351
8	4	3	8	45	25							13.8		2		30	20	17.5							
8	4	3	8	43	25							14.3		0		30	20	8							
8	4	3	8	45	25							13.8		2		30	20	17.5							
8	4	3	8	45	25							13.5		2+1		30	20	17.5							
8	4	3	8	36	25							3.2		0		30	20	8							
8	4	3	8	45	25							14.8		2+1+1		30	20	17.5+4							
8	4	3	8	45	25							12		2+3		30	20	17.5							
8	4	3	8	45	25							14.3		2+2+1		30	20	18+4		184.3	2723	2918	1080	540	7461
8	4	3	8	45	25							15		2		30	20	8+4		172	2729		360		3089
8	4	3	8	45	25							10.4		2		30	20	17.5		174.9	2775		360		3135
8	4	3	8	45	25							11.9		2		30	20	17.5		174.4	2767		360		3127

Only variation in:
-Attendance
-`Management'
-Papers written

Rest of the Presentation

- Background/Context
  - Study in Rural China
  - Teacher Performance Pay Policy in 2009
- **Experimental Design and Interventions**
- Sampling/Data
- Results + Discussion

Experimental Design

Teacher incentive treatment (outcome-based design feature x payout size design feature)	X. Large incentive payout	Y. Small incentive payout
A. Control	A. 52 schools	
B. Levels incentive	BX. 26 schools	BY. 28 schools
C. Gains incentive	CX. 26 schools	CY. 30 schools
D. Pay for percentile incentive	DX. 26 schools	DY. 28 schools

Math teachers in 216 schools

Approximately 8,000 grade 6 students

Underlying Structure
(Common to all treatment groups)

- Incentives tied to student achievement as measured by scores on standardized math exams

- Teachers compete in rank-order tournament with teachers in other schools

- No explicit penalty for missing students, but potential disqualification

Design Feature 1: Different ways of using student achievement to measure and reward teacher performance (Teacher Performance Indices)

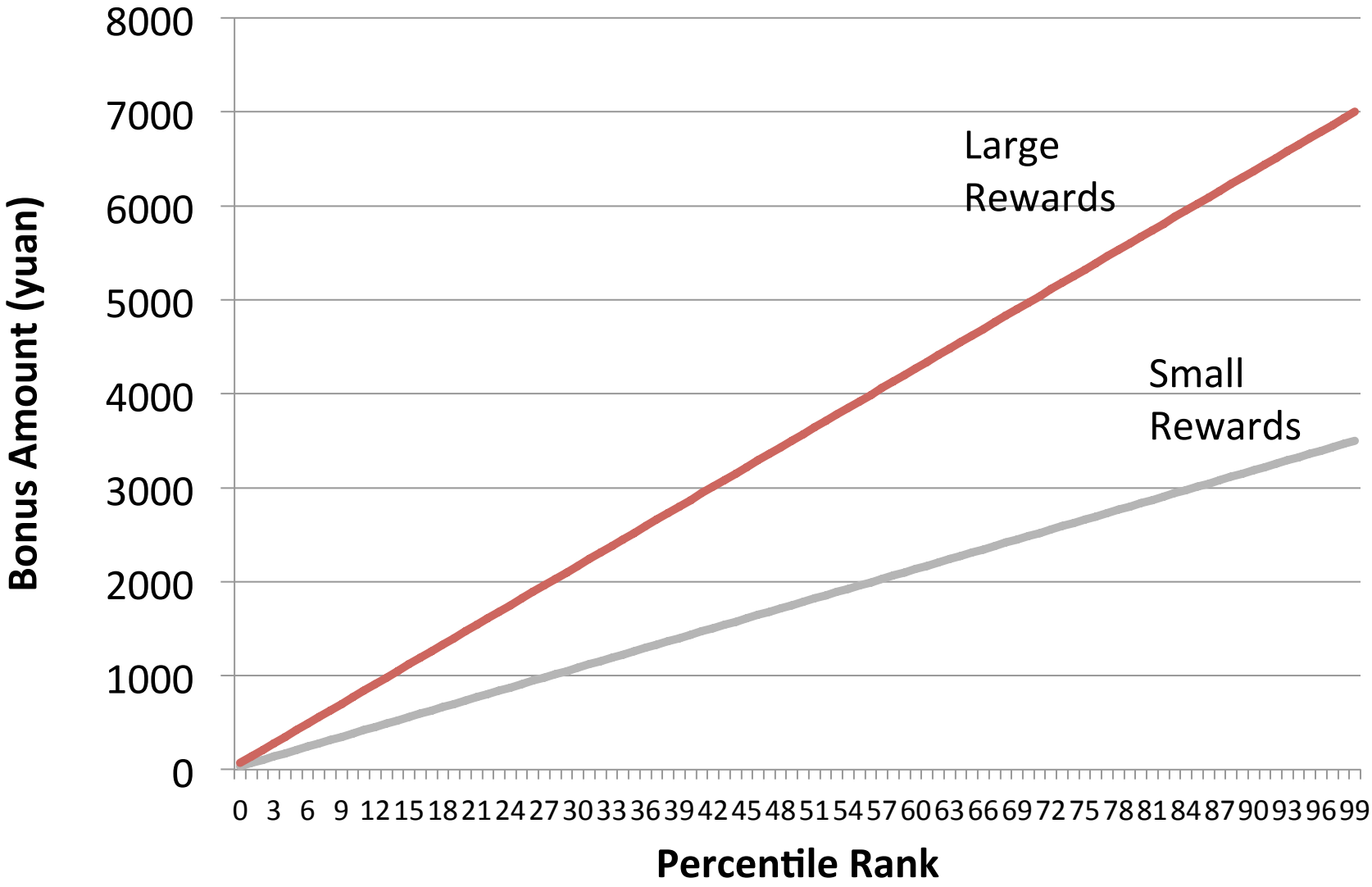
Levels Incentive: Rewards teachers based on student achievement on an end-of-the-year exam

Gains Incentive: Rewards teachers based on gains in achievement from the start to the end of the year

Pay for percentile incentive: Reward teachers based on pay for percentile index: Within similar comparison sets (among students with similar baseline scores), rank students by scores on endline exam and give them a percentile rank. Averaged them to create pay for percentile index (Neal, 2011).

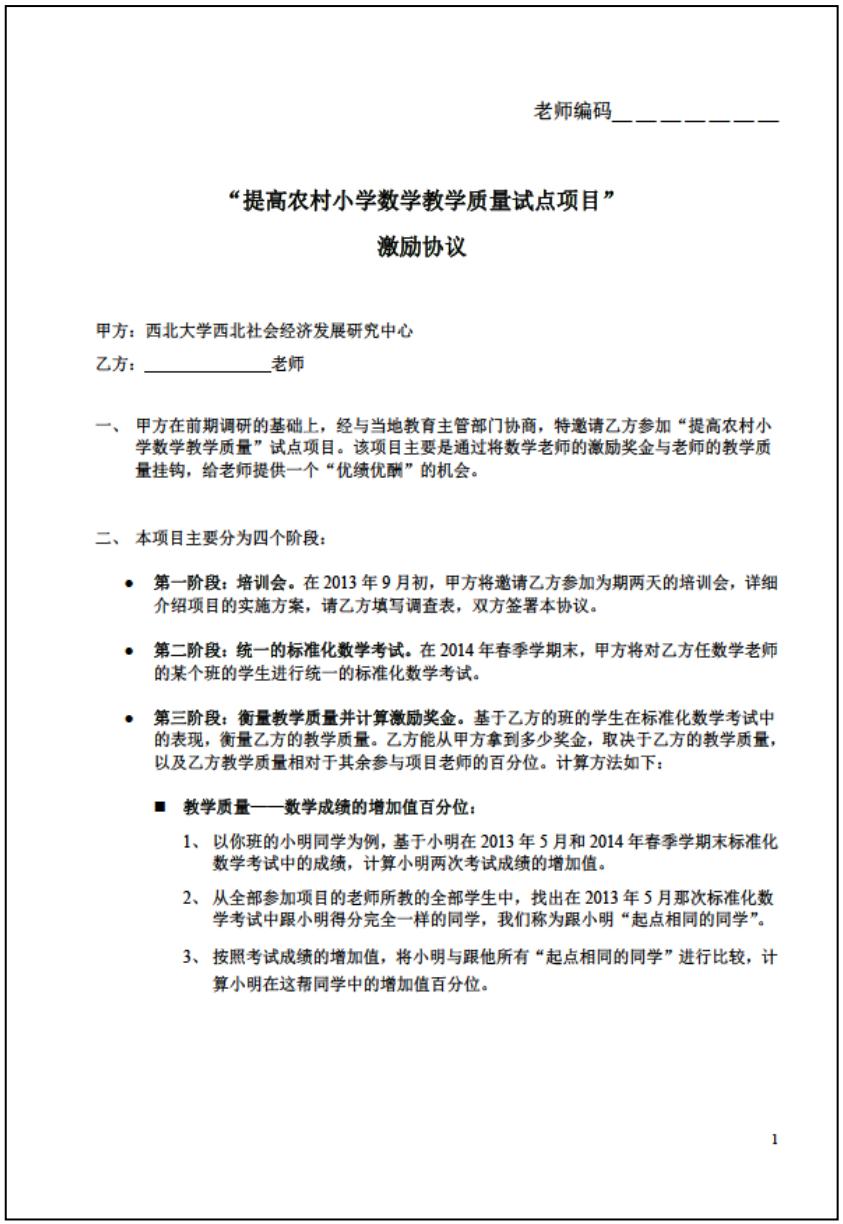
- Explicitly accounts for multiple students (Barlevy & Neal, 2012)

Design Feature 2: Large vs. Small Rewards



Top reward in small group ≈ 1 month pay

Incentive “Agreement”



老师编码_ _ _ _ _ _ _ _

“提高农村小学数学教学质量试点项目”

激励协议

甲方：西北大学西北社会经济发展研究中心

乙方：_____老师

一、 甲方在前期调研的基础上，经与当地教育主管部门协商，特邀请乙方参加“提高农村小学数学教学质量”试点项目。该项目主要是通过将数学老师的激励奖金与老师的教学质量挂钩，给老师提供一个“优绩优酬”的机会。

二、 本项目主要分为四个阶段：

- **第一阶段：培训会。**在 2013 年 9 月初，甲方将邀请乙方参加为期两天的培训会，详细介绍项目的实施方案，请乙方填写调查表，双方签署本协议。
- **第二阶段：统一的标准化数学考试。**在 2014 年春季学期末，甲方将对乙方任数学老师的某个班的学生进行统一的标准化数学考试。
- **第三阶段：衡量教学质量并计算激励奖金。**基于乙方的班的学生在标准化数学考试中的表现，衡量乙方的教学质量。乙方能从甲方拿到多少奖金，取决于乙方的教学质量，以及乙方教学质量相对于其余参与项目老师的百分位。计算方法如下：
  - **教学质量——数学成绩的增加值百分位：**
    1、 以你班的小明同学为例，基于小明在 2013 年 5 月和 2014 年春季学期末标准化数学考试中的成绩，计算小明两次考试成绩的增加值。
    2、 从全部参加项目的老师所教的全部学生中，找出在 2013 年 5 月那次标准化数学考试中跟小明得分完全一样的同学，我们称为跟小明“起点相同的同学”。
    3、 按照考试成绩的增加值，将小明与跟他所有“起点相同的同学”进行比较，计算小明在这帮同学中的增加值百分位。

1

Incentive Agreement Guide

Rest of the Presentation

- Background/Context
  - Study in Rural China
  - Teacher Performance Pay Policy in 2009
- Experimental Design and Intervention
- **Sampling/Data**
- Results + Discussion

Sampling and Data

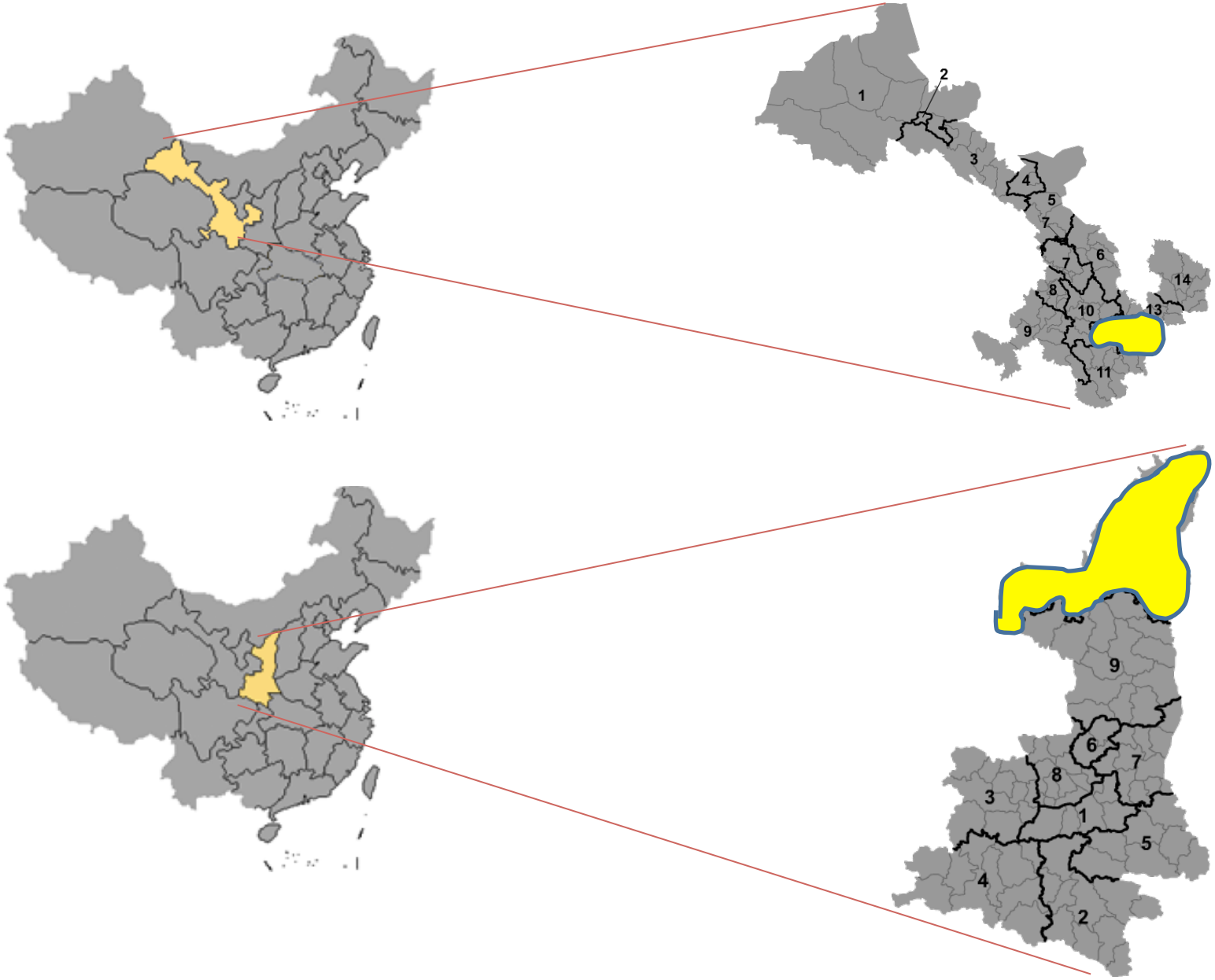
Sample

- 16 Counties in Tianshui (Gansu) and Yulin (Shaanxi)
- 216 Schools (243 Math Teachers)
- All 6th grade students, about 8,000 Students Total

Data

- 2 waves of pre-program math scores
- Teacher Survey at Baseline (Sept. 2013)
  - Detailed information on teacher characteristics, existing incentives, perceptions, social preferences
- Endline Math Exam (constructed test w/ good properties)
- Detailed Student, Teacher, School Surveys (May 2014)

Sampling: Study Locations



Estimation Strategy

- Main specification (for child *i* in school *j*):

$$Y_{ijc} = \alpha + T'_{jc}\beta + x'_{ijc}\Upsilon + \lambda_c + \varepsilon_{ijc}$$

  - Y_{ijc} Outcome of interest at the endline (e.g. math scores)
  - T_{jc} Vector of treatment dummies
  - x_{ijc} Baseline student, teacher, school characteristics
  - λ_c Block/strata (county) fixed effects
  - Standard errors account for clustering at the school level

- Significance based on p-values adjusted for multiple hypotheses (Romano and Wolf)
- *Pre-analysis plan filed in AEA registry before data available*

- *Balance across treatment arms on baseline characteristics*

Rest of the Presentation

- Background/Context
  - Study in Rural China
  - Teacher Performance Pay Policy in 2009
- Experimental Design and Intervention
- Sampling/Data
- **Results + Discussion**

Average Impacts on Math Scores
(Design Feature 1: Teacher Performance Indices)

	(1)	(2)	(3)	(4)
Levels	0.056	0.084		
	(0.048)	(0.052)		
Gains	0.012	0.001		
	(0.051)	(0.050)		
Pay-for-percentile	0.128*	0.148**		
	(0.064)	(0.064)		
Small			0.063	0.081
			(0.053)	(0.055)
Large			0.064	0.067
			(0.045)	(0.046)
Baseline Scores	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes
P-value: Gains - Levels	0.390	0.114		
P-value: P4Pct - Levels	0.236	0.292		
P-value: P4Pct – Gains	0.078	0.023**		
P-value: Large – Small			0.989	0.778
Observations	7,454	7,373	7,454	7,373

Robust standard errors accounting for clustering at the school level in parentheses. ** p<0.05, * p<0.1 after adjustment.

Average Impacts on Math Scores
(Design Feature 2: Large vs Small Rewards)

	(1)	(2)	(3)	(4)
Levels	0.056	0.084		
	(0.048)	(0.052)		
Gains	0.012	0.001		
	(0.051)	(0.050)		
Pay-for-percentile	0.128*	0.148**		
	(0.064)	(0.064)		
Small			0.063	0.081
			(0.053)	(0.055)
Large			0.064	0.067
			(0.045)	(0.046)
Baseline Scores	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes
P-value: Gains - Levels	0.390	0.114		
P-value: P4Pct - Levels	0.236	0.292		
P-value: P4Pct – Gains	0.078	0.023**		
P-value: Large – Small			0.989	0.778
Observations	7,454	7,373	7,454	7,373

Robust standard errors accounting for clustering at the school level in parentheses. ** p<0.05, * p<0.1 after adjustment.

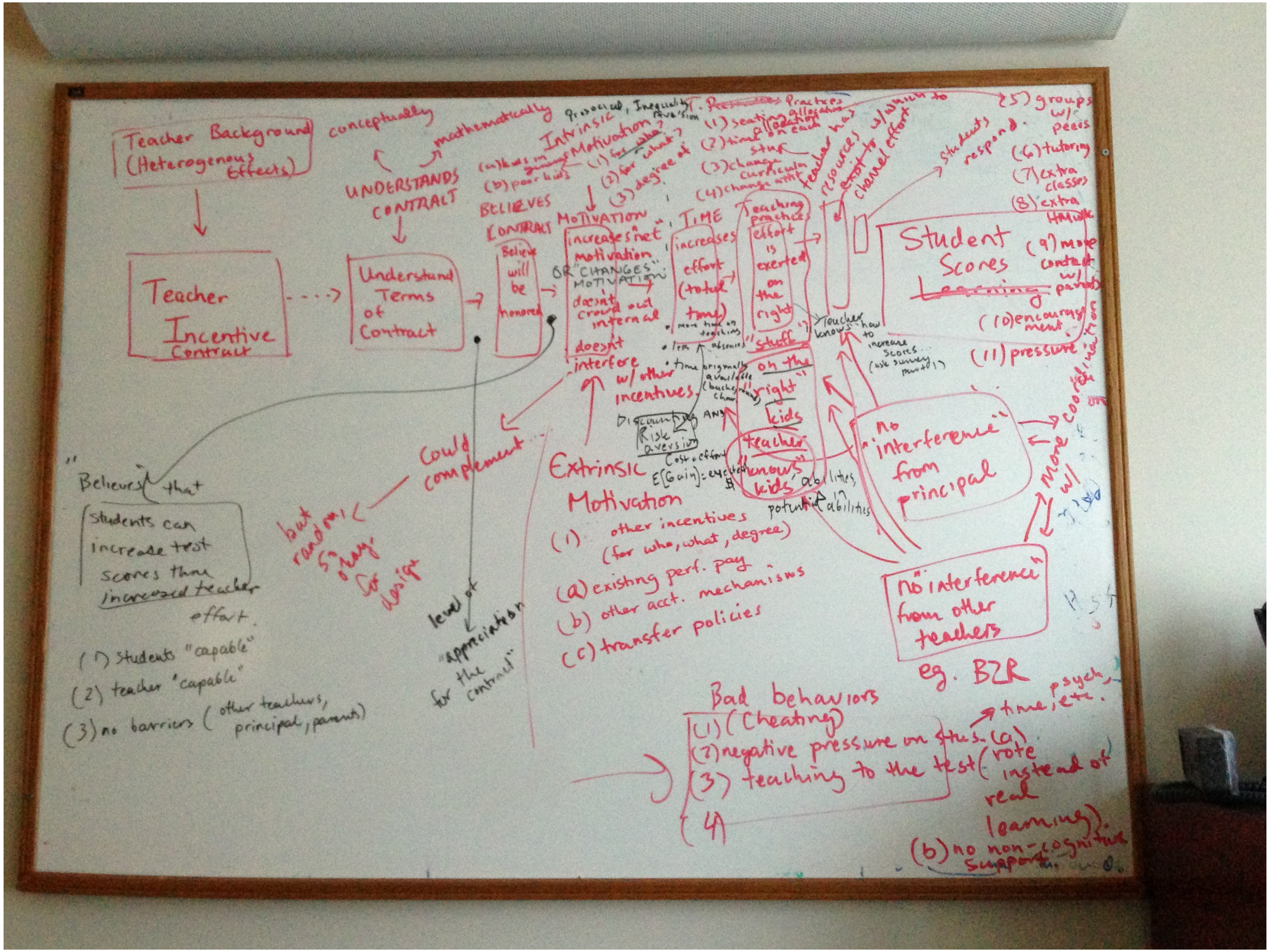
Average Impacts on Math Scores
(Teacher Performance Index by Reward Size)

Note: Not Pre-specified

	Small		Large	
	(1)	(2)	(3)	(4)
Levels	0.046	0.080	0.064	0.081
	(0.059)	(0.067)	(0.059)	(0.061)
Gains	0.049	0.037	-0.033	-0.033
	(0.064)	(0.063)	(0.060)	(0.061)
Pay-for-percentile	0.089	0.131	0.163**	0.165**
	(0.094)	(0.100)	(0.059)	(0.060)
Baseline Scores	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Other Controls		Yes		Yes
P-value: Gains - Levels	0.974	0.730	0.153	0.100
P-value: P4Pct - Levels	0.648	0.667	0.157	0.237
P-value: P4Pct - Gains	0.690	0.546	0.005**	0.004**
Observations	4655	4609	4678	4628

Robust standard errors accounting for clustering at the school level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Other controls include student gender, age, parent's education, a household wealth index, class size, teacher experience and teacher base salary.

Mechanisms


Teacher Background (Heterogeneous Effects)
Teacher Incentive Contract
Understand Terms of Contract
UNDERSTANDS CONTRACT
BELIEVES CONTRACT
Believe will be honored
Intrinsic Motivation
MOTIVATION
increases net motivation
TIME
increases effort (total time)
Teaching practices effort is exerted on the right "stuff" on the "right" kids
Student Scores
Extrinsic Motivation
(1) other incentives (for who, what, degree)
(a) existing perf. pay
(b) other acct. mechanisms
(c) transfer policies
Could complement
"no interference" from principal
"no interference" from other teachers
eg. B2R
Bad behaviors
(1) (Cheating)
(2) negative pressure on stus.
(3) teaching to the test
(4)
(5) groups w/ peers
(6) tutoring
(7) extra classes
(8) extra HMwk
(9) more contact w/ parents
(10) encouragement
(11) pressure
Believes that students can increase test scores thru increased teacher effort.
(1) students "capable"
(2) teacher "capable"
(3) no barriers (other teachers, principal, parents)
level of "appreciation" for the contract

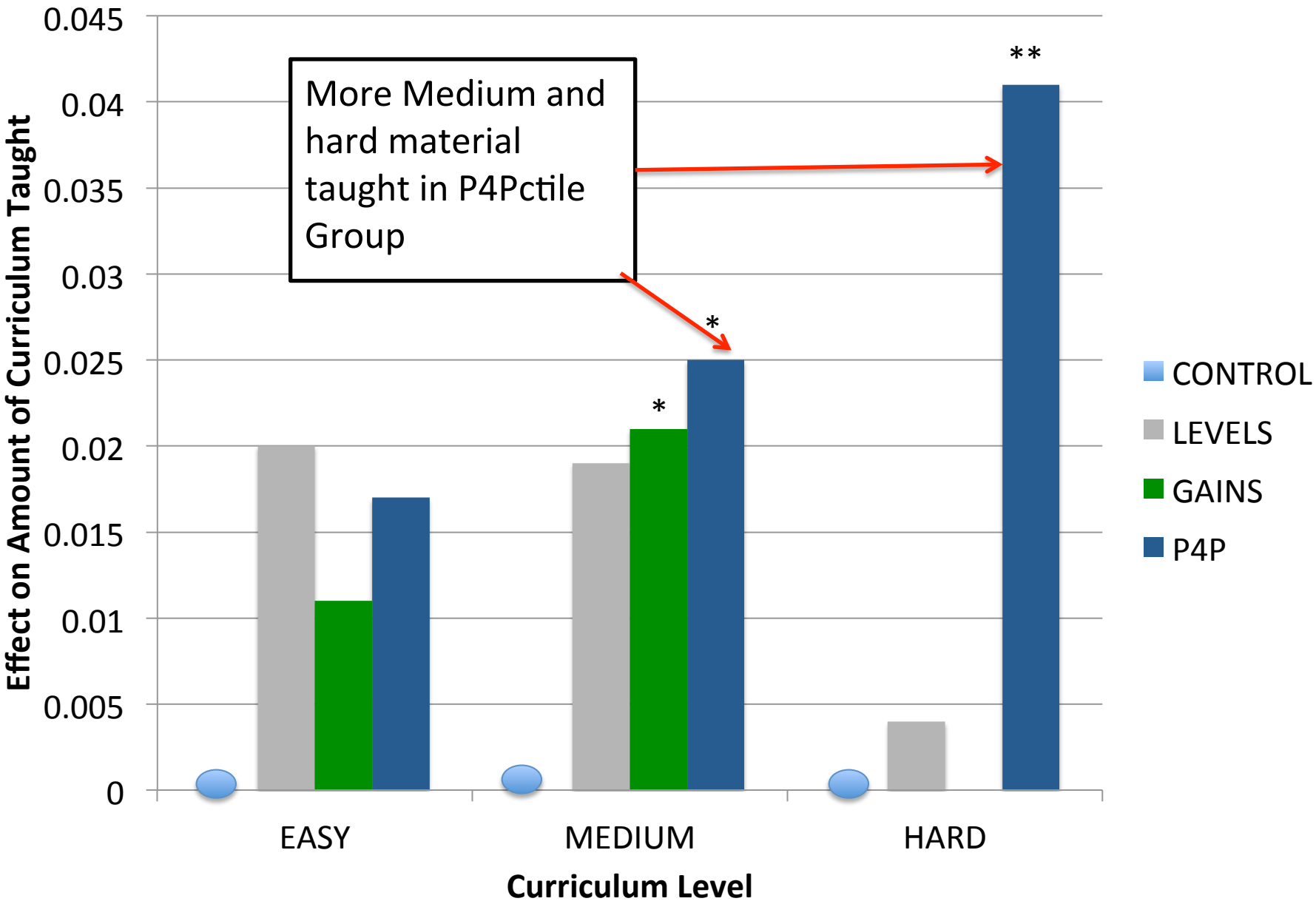
Impacts on teacher behavior

- Like Glewwe et al. (2010) and Muralidharan and Sundararaman (2011), we find little effect on many types of teacher behavior in the classroom (reported by students):
  - Classroom engagement
  - Care
  - Classroom management
  - Communication with students

- We find no significant effect on self-reported teacher effort.

- While we do find impacts of *all types* of incentives on student-reported times being tutored outside of class, these do not explain the significantly larger impact of pay-for-percentile.

Effect on Amount and Type of Curricula Taught



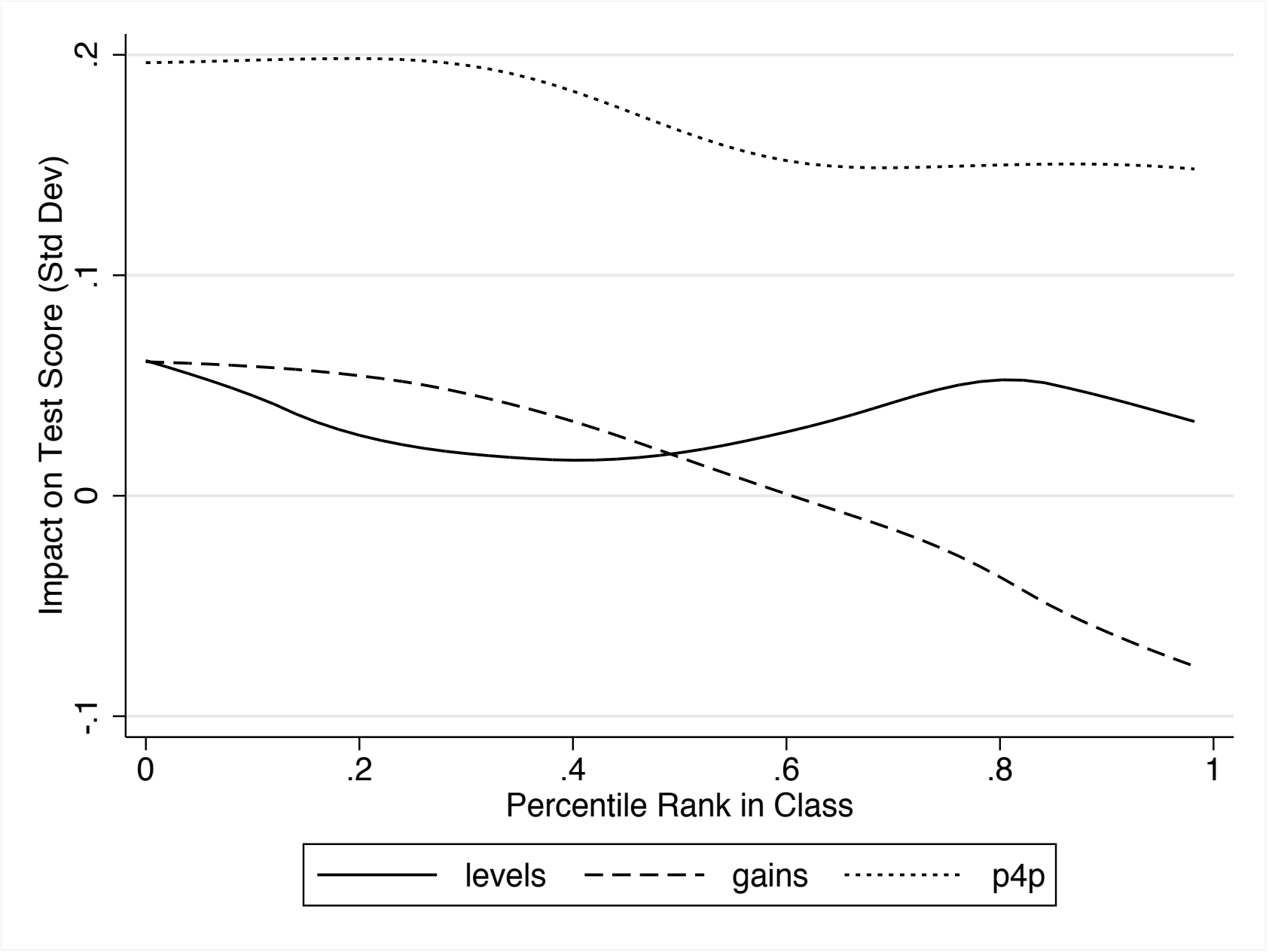
More curricula coverage – was it at the expense of intensity of instruction (teachers just went faster)?

Treatment effects (large rewards group) on easy, hard, medium test items:

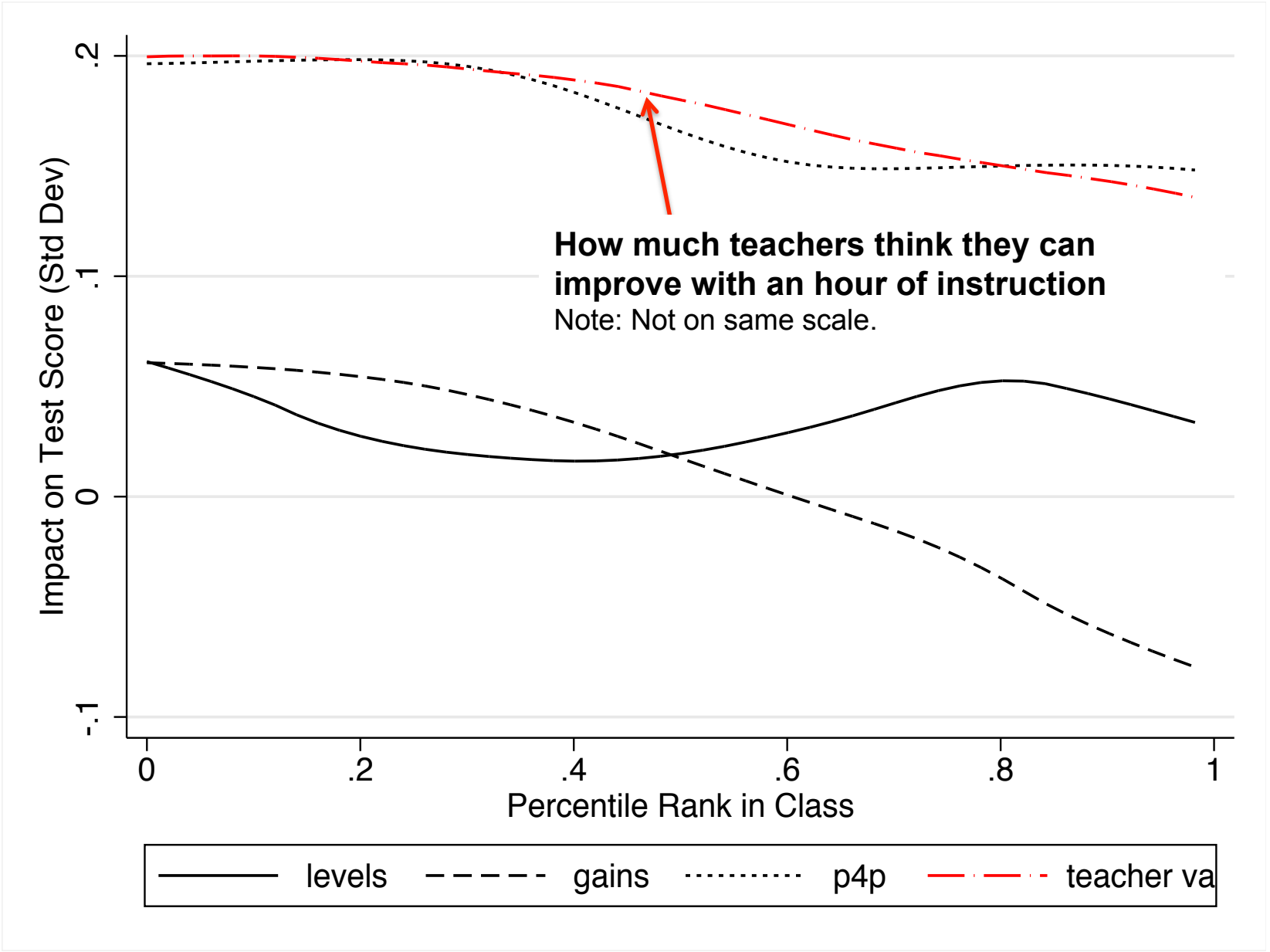
		Easy (1)	Medium (2)	Hard (3)
(1)	Levels Incentive	0.029	0.094	0.075
		(0.044)	(0.50)	(0.052)
(2)	Gains Incentive	-0.006	-0.010	0.019
		(0.036)	(0.050)	(0.053)
(3)	Pay-for-Percentile Incentive	0.105**	0.092	0.16**
		(0.043)	(0.062)	(0.067)

Pay-for-percentile led to gains in easy & hard items– suggesting it increases both the coverage & intensity of instruction

Distributional Effects of the Incentive Designs: Within Classes



Distributional Effects of the Incentive Designs: Within Classes



Summary of Results

- Of the different teacher performance indices used to incentivize teachers, only pay for percentile has significant effects on average. It is accompanied by meaningful changes in curricular coverage

- Doubling size of reward does not have statistically significant effect
  - limited power to test within incentive designs
  - Only large increase in point estimate for pay-for-percentile

- Pay for percentile produced broad-based gains across distribution of student achievement
  - For levels, gains teachers focus on students for whom perceived returns to effort highest

Limitations/Contributions

Limitations: Only examined impacts after year 1.

Contributions:

1) Head-to-head experimental test of alternative approaches of mapping student achievement into rewards for individual teachers (including "pay-for-percentile")
   - Adds to previous work testing individual vs group incentives (Muralidharan and Sundararaman 2011; Behrman et al. 2012), standard vs loss aversion-based incentives (Fryer et al. 2012)
2) First experimental study of reward size in teacher performance pay
3) Closely examine how different performance pay designs affect distribution of student achievement within the class (and how teachers likely multitask across students)

THANK YOU!

ssylvia@ruc.edu.cn

reap.stanford.edu