

BREAD-IGC Virtual Development Course

Education: Technology in Education

Karthik Muralidharan

UC San Diego, NBER, BREAD, and J-PAL

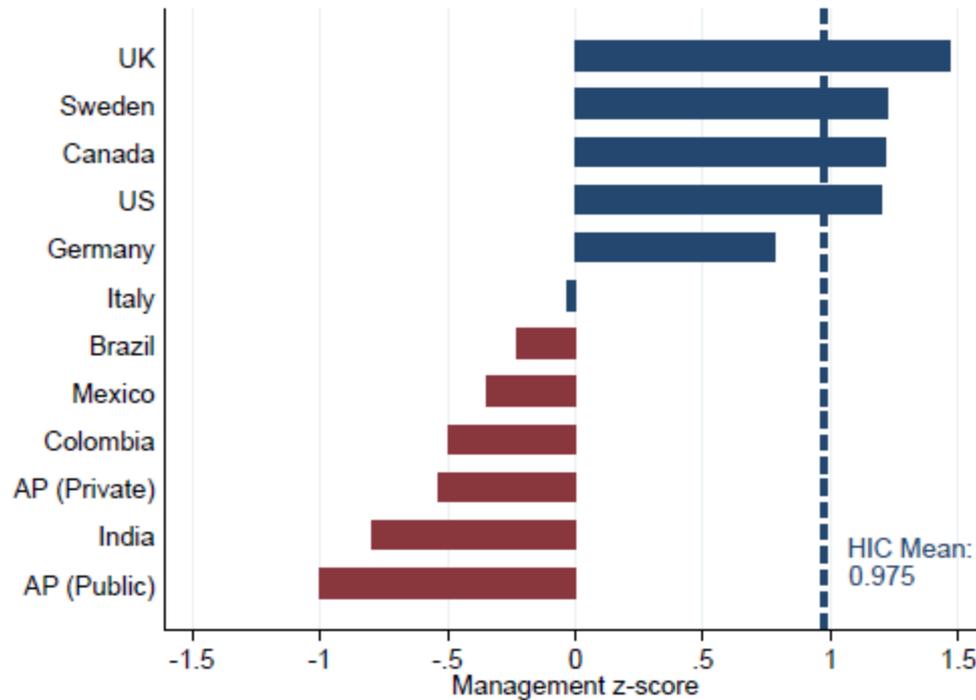
11 February 2022

School Governance Wrap-up

- Teacher absence
- Monitoring: top-down & bottom-up
- Teacher pay and teacher incentives
- Contract teachers
- **School management**

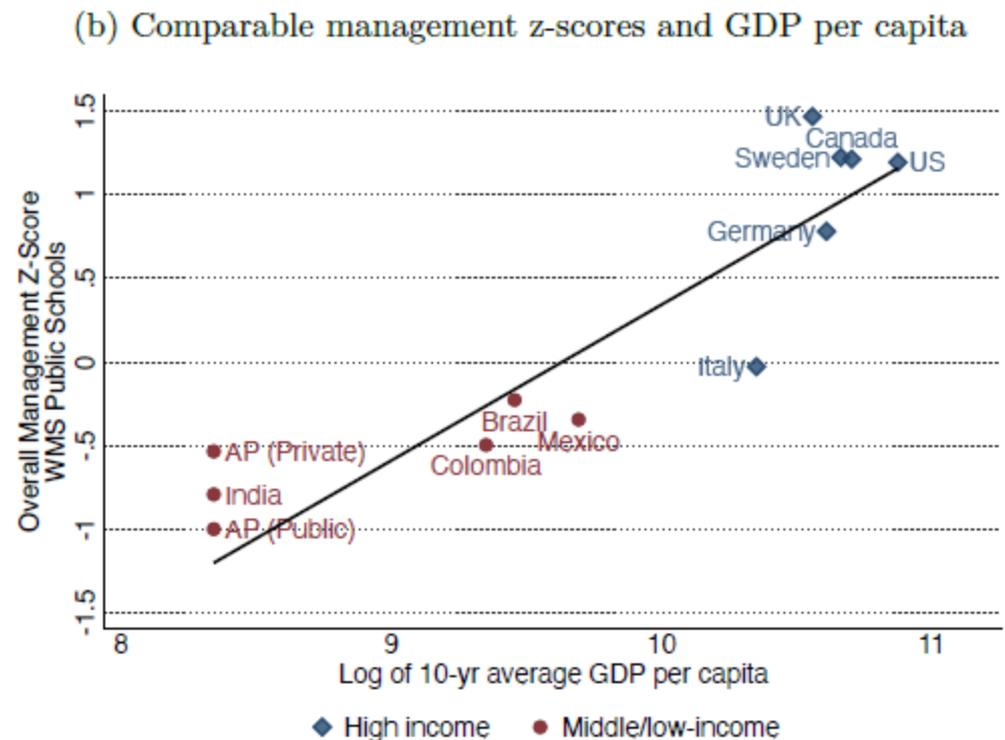
Figure 1: Global benchmarks

(a) Rank of comparable management z-scores

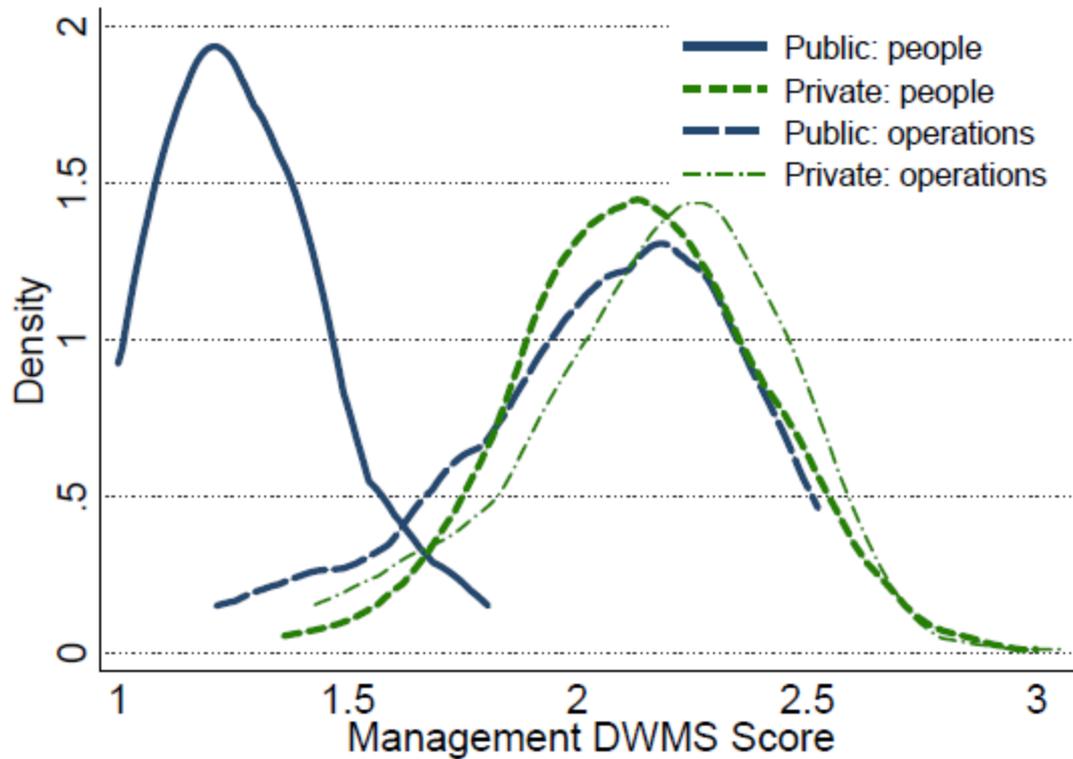


an, and
r(2021)

- Develop and use the D-WMS to be comparable to the WMS but much more granular to allow us to better capture finer variation in the left tail
- School management scores are systematically lower in low-income countries



- Not an outlier after adjusting for income – but suggests that variation in TFP in education systems may also partly explain income-outcome gradient
- Thus, improving quality of school management may be an important component of improving developing country education systems



- Public schools do especially poorly on personnel management
- Consistent with differences in absence, active teaching, etc. observed in the data
- Also consistent with our finding a strong positive correlation between teacher value-added and pay in private schools, and no such correlation in public schools
- Management scores are also correlated with teaching activity & value added

Can we improve public school management at scale?

- LMIC Governments increasingly understand the importance of school management and are keen to improve it
 - Many such reforms (160 examples in 84 countries in WB database)
 - But very little evidence on impact
- Muralidharan & Singh (2021): Large-scale RCT of a school management intervention in the Indian state of MP
- The program (MPSQA) was modelled after several global ‘best practices’ in school management (and developed with several international expert inputs)
 - Comprehensive school quality assessments
 - Customized school improvement plans
 - Regular (intended) follow up by cluster-resource coordinators
 - However, no incentives (explicit or implicit)
- Randomized and evaluated at scale with 1,774 treatment schools

Results

Four main results:

1. Assessments were completed and of

- ▶ ~93% of elementary schools assigned were covered by the assessments and had S
- ▶ Contain meaningful variation across sc
- ▶ Most schools rated inadequate (collusi

2. No improvement in support or oversi

- ▶ No change in frequency of visits or the
- ▶ School Management Committees also
in treatment schools

3. No change in teacher effort and clas

Evaluating a further scale-up

- ▶ The government had already planned for the program to the next phase of $\sim 25,000$ schools
- ▶ Additional impetus from a similar nation
 - ▶ External assessments were replaced by internal ones
 - ▶ The plans were made much more detailed
 - ▶ No change in incentives
 - ▶ Already scaled up to $>600k$ schools, to $>100k$ schools
- ▶ We evaluate this scale up using a matched control design

Understanding program implementation

- ▶ We collected extensive qualitative data from teachers and education officials
- ▶ We document that, for these officials, the reform was an exercise in administrative compliance
 - ▶ Both teachers and supervisors perceived the reform as a data collection and paperwork filling exercise.
 - ▶ Paperwork was submitted on time.
 - ▶ Program delivery effectively ceased after the first year.
 - ▶ De facto, the reform was very far from what was intended.

Contributions 1/2

Improving management quality in developing countries

- ▶ Improving management quality in developing countries
 - ▶ Management quality is correlated with growth in both the private and public sectors (Bloom and Ross 2018, Rasul et al 2018)
 - ▶ Providing management consulting improved productivity in private firms (Bloom et al 2013, Bruhn et al 2018)
- ▶ We provide experimental evidence on how to improve management quality in the **public** sector
 - ▶ No impact on either processes or outcomes
 - ▶ Consistent with other RCT evidence that

Contributions 2/2

Organizational Economics and Public Sector Reform

- ▶ Shows the difficulties of “change management”
 - ▶ Large and active literature, with several studies but very little well-identified evidence
 - ▶ Importance of well-identified null results
 - ▶ Null results may also reflect intensity of effort
- ▶ Illustrate the nature of bureaucratic incentives (the gap between actual and perceived ‘success’)
 - ▶ The program was deemed a success because of the perceived success
 - ▶ We show that the program’s success was largely due to the perceived success

Bureaucratic incentives under weak s

*What stands out here are higher-level
administrative hierarchy making decisions about
that bear little relevance to realities on t
in turn, are **subordinates faithfully ex**
paper but caring little for how well t
Targets are indeed met, but the ultimate
go unfulfilled.*

Why do governments continue scaling

Institutional isomorphism

One compelling explanation is provided by the concept of "institutional isomorphism"

"Organizations tend to model themselves after other organizations in their field that they perceive to be successful," (p. 152);

"these institutional isomorphic processes tend to proceed [even] in the absence of evidence of organisational efficiency," (p.153);

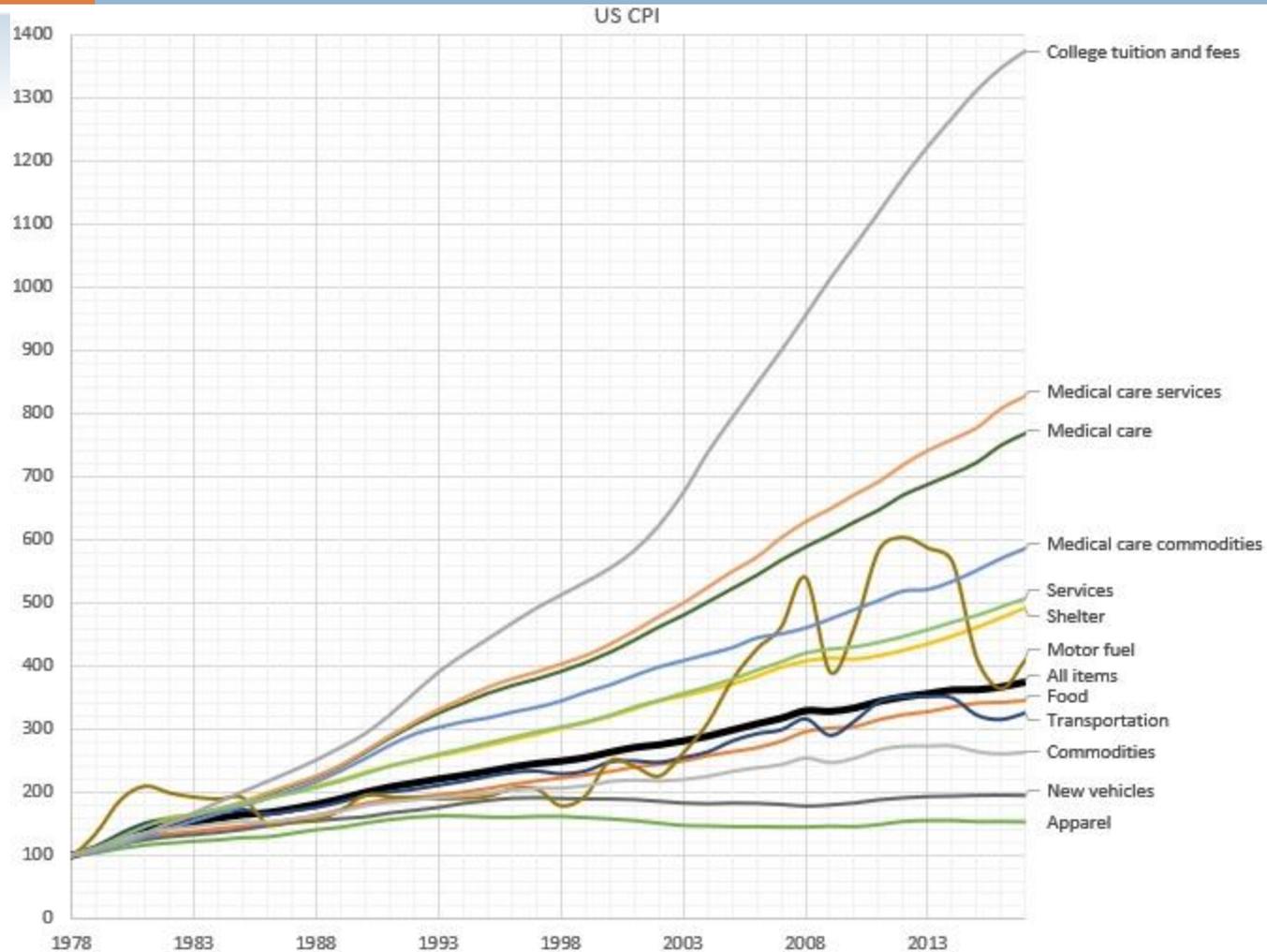
Conclusions - What should we take away from all of this?

- Returns to improving management quality may be especially high in public sector
 - But it may also be *much* more difficult given nature of bureaucratic incentives
- We cannot identify what factors *would* lead to success
- But 3 factors seem quite important (from other studies of interventions that have been found to work)
 - Better incentives for improving effort or outcomes
 - Better visibility on outcomes at the beneficiary level
 - Additional staffing
- Governments and donors are constantly designing and deploying programs to improve service delivery and development outcomes
 - Programs are often judged on design quality (“best practices”) and number of people reached
 - On these measures, the program was a resounding success!
 - Highlights importance of independent evaluations of impact

Technology in Education

- Productivity gains in education production have been limited compared to rest of the economy (around the world)
 - Seen in higher cost inflation relative to CPI
 - Case of Baumol's disease – education & health (education may be worse)
 - Technology of instruction has basically been unchanged (chalk & talk)
 - Hence, lots of excitement about technology in education/instruction
- Mechanisms of potential impact include
 - Cost-effective access to high-quality instruction
 - Leapfrog constraints in teacher knowledge
 - Supplemental instruction, practice, reinforcement at home
 - Customizing learning paths for students (and inducing greater engagement)
 - Shortening feedback loop
 - Better engagement with parents
- Technology can also be used to improve governance include
 - Monitoring teacher attendance; improving outcome measurement

Baumol's Disease Illustrated



However evidence on CAL has been mixed & not lived up to hype

- Banerjee et al (2007)
 - Large positive effects (0.47SD) of a CAL program run by Pratham
 - However, found to be 5-7 times less cost effective than using balsakhis (similar demographic as the contract teachers)
- Cristia et al (2017) & Beuermann et al (2015)
 - RCT of the impact of school and home use of the XO Laptop (OLPC)
 - Positive impacts on knowledge of computer use
 - But no impact on measures of cognitive development
- Malamud & Pop-Eleches (2011)
 - RD-based study of providing home computers to middle-schoolers in Romania
 - Improved computer skills
 - But *hurt* grades on core school subjects (Math, English, Romanian)
 - Most students report playing computer games on a daily basis (!); suggestive evidence of reduced time reading and doing homework
- Overall, the evidence points to: “mixed evidence with a pattern of null effects” (Bulman and Fairlie 2016 review article)

The *Mindspark* program

- ▶ We present an experimental evaluation of the program (*Mindspark*) in India
 - ▶ Developed by Educational Initiatives (EduTech), a leading educational assessments provider, over a 10-year period
 - ▶ Over 45,000 question Item Bank, used for administering over a million questions
 - ▶ Provides individual, dynamically updated instruction
 - ▶ Instruction is targeted at children's actual learning level **the curriculum-mandated level**
 - ▶ The software itself is platform-agnostic, cloud-based, internet-based, and after-school model

Summary of results

- ▶ **Business-as-usual learning:**

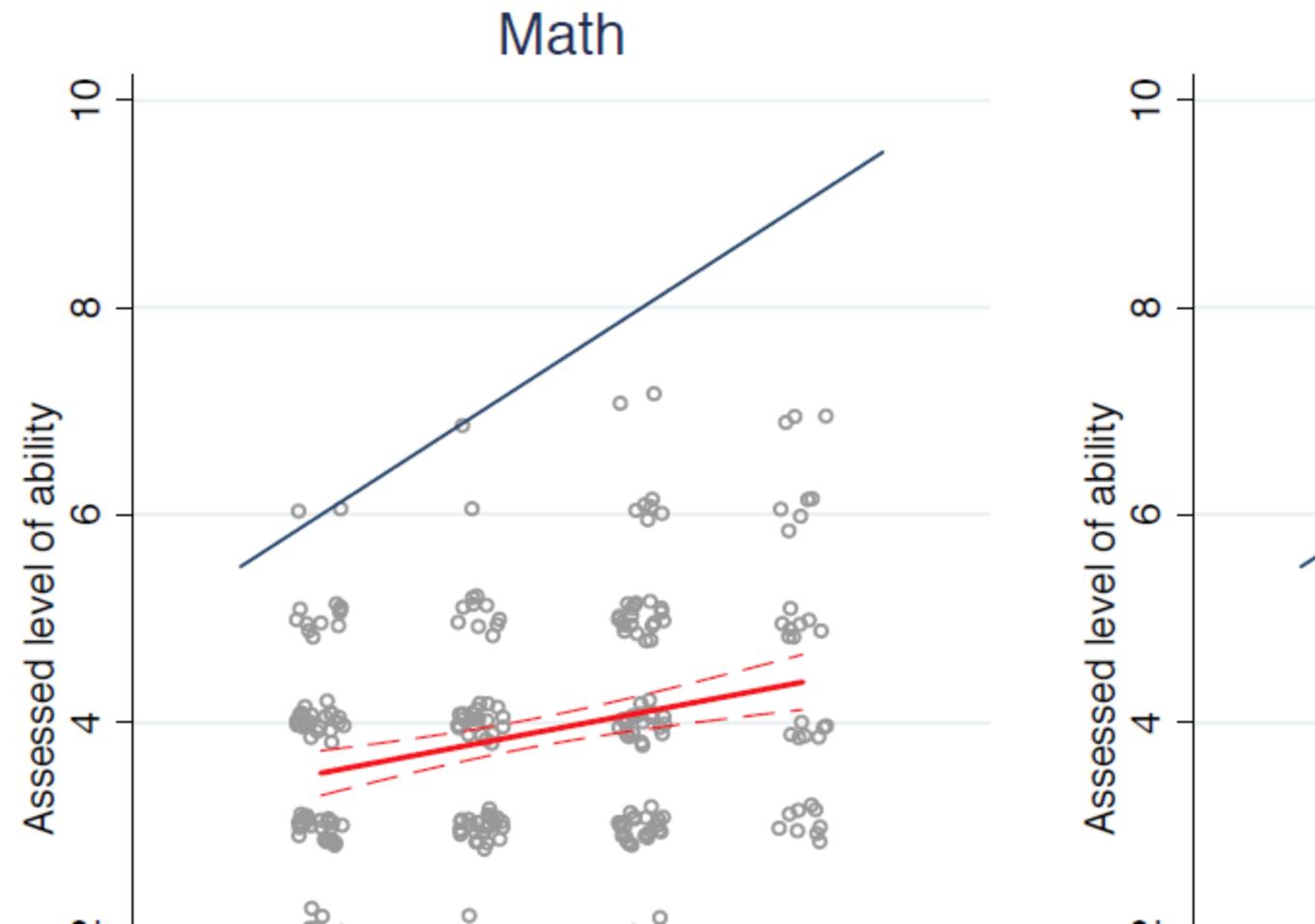
- ▶ Students in this setting are several grades below the grade-appropriate standard and this gap is persistent
- ▶ In the control group, students in the bottom 25% of the within-grade distribution show zero progress over the year on our independently-administered tests

- ▶ **Program effects:**

- ▶ The offer of a Mindspark voucher led to a 0.22 SD increase in math and 0.22 SD in Hindi over the study period
 - ▶ IV estimates: 0.59 SD/0.36 SD in math and 0.48 SD/0.36 SD in Hindi
- ▶ Large and similar absolute test-score gains in math and Hindi

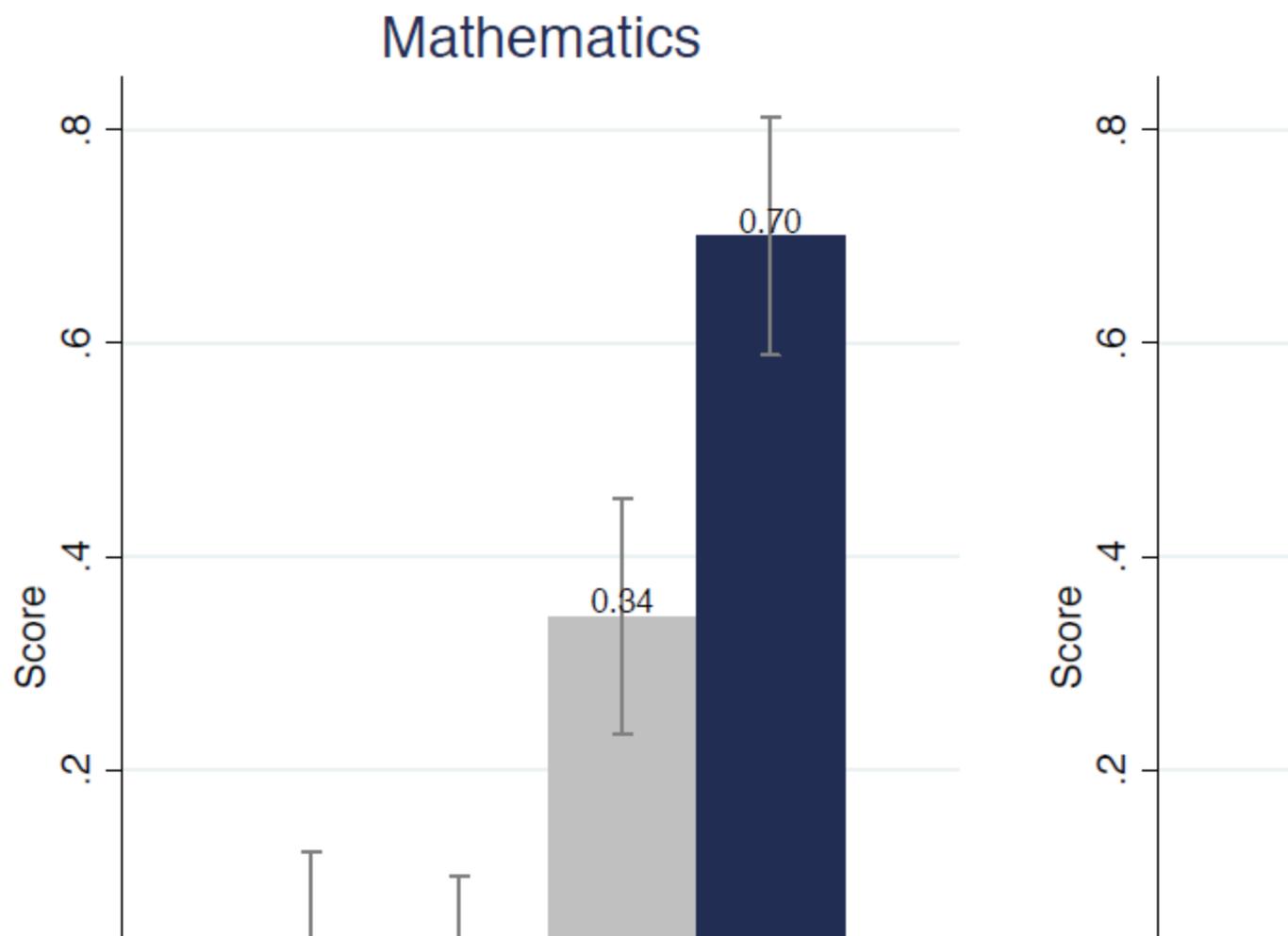
Actual vs. expected learning levels

In the treatment group at start of intervention



The core result

Mean differences in achievement

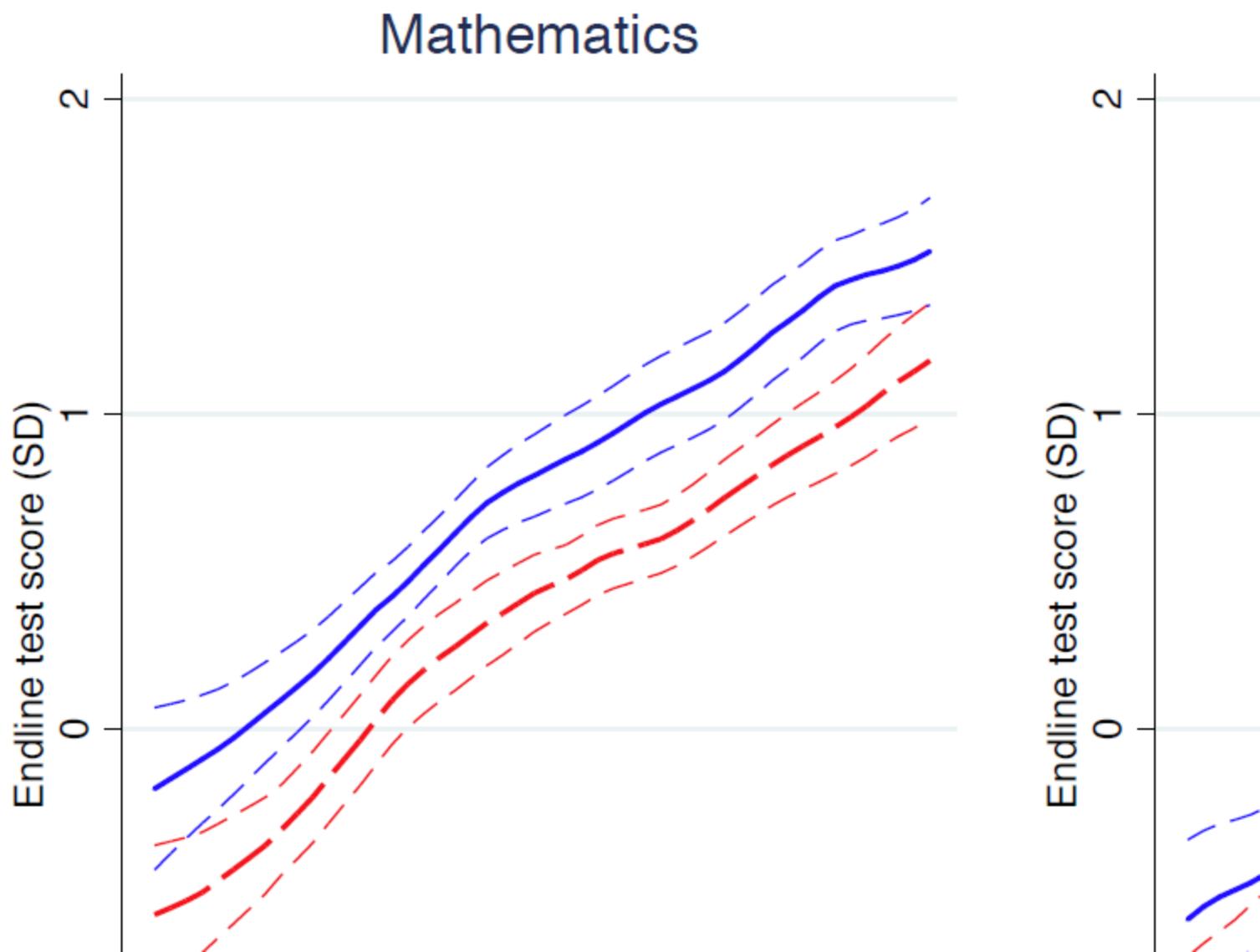


The core result

ITT Results in a regression framework

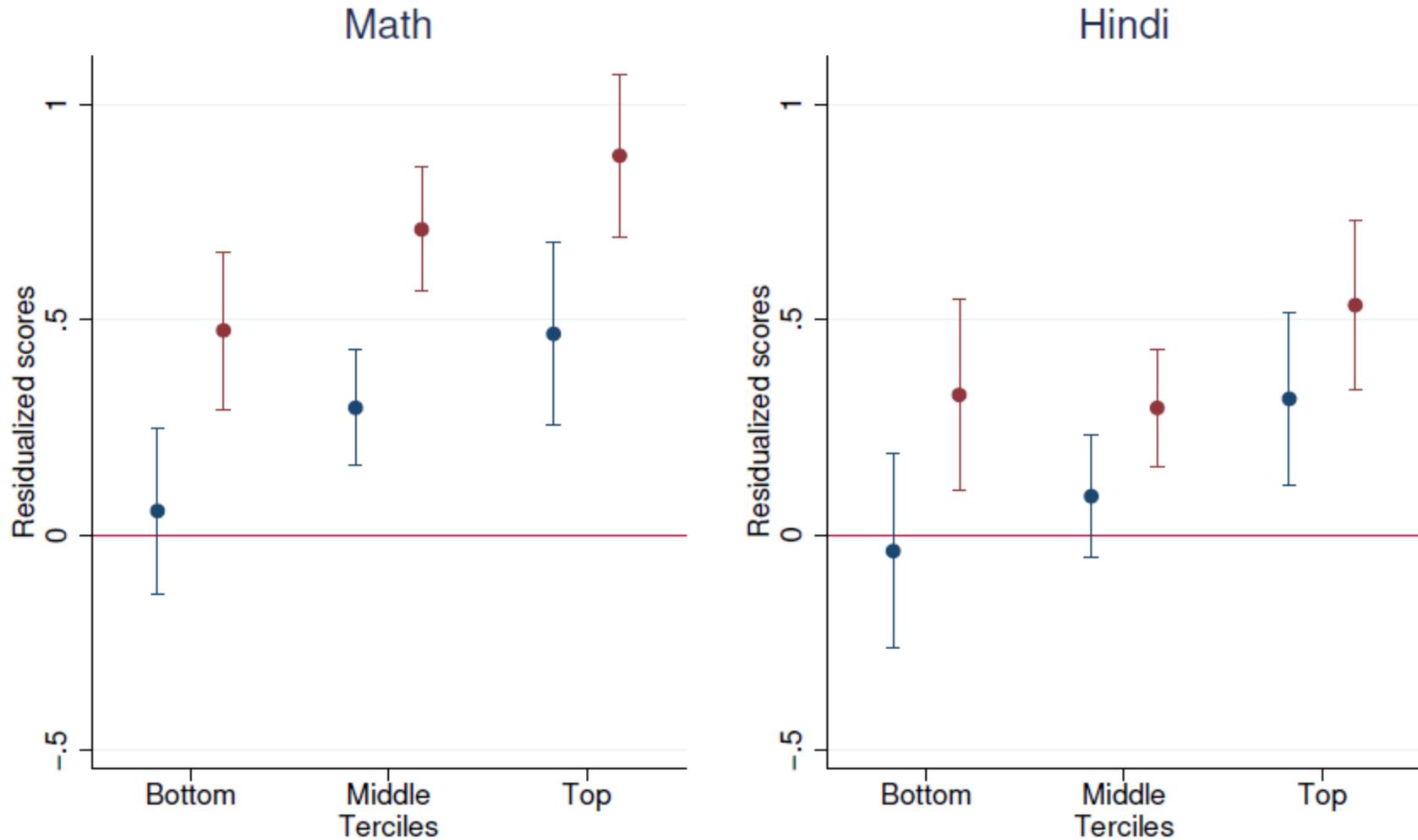
	(1)	(2)
	Dep var: Math	Standardize Hindi
Treatment	0.36*** (0.063)	0.22*** (0.076)
Baseline score	0.54*** (0.047)	0.67*** (0.034)
Constant	0.36*** (0.031)	0.15*** (0.038)

Learning gains across the full distrib



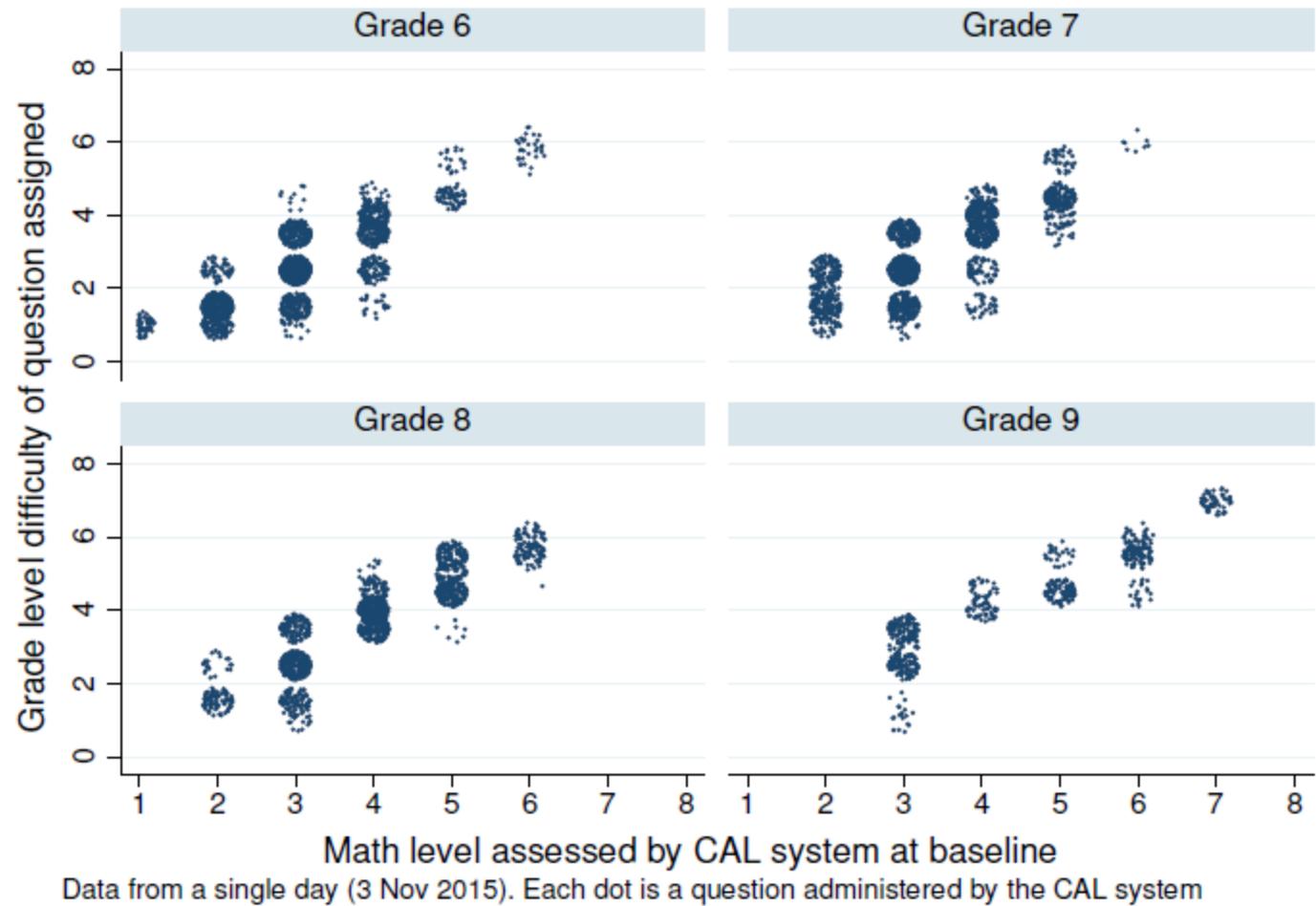
Treatment vs. “business-as-usual” progress

Children in the lowest terciles make *zero progress* in control



No single teacher can individualize instruction so finely

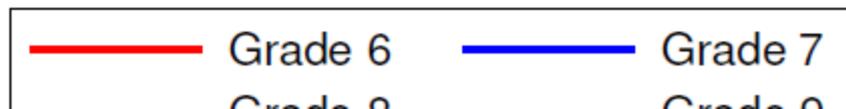
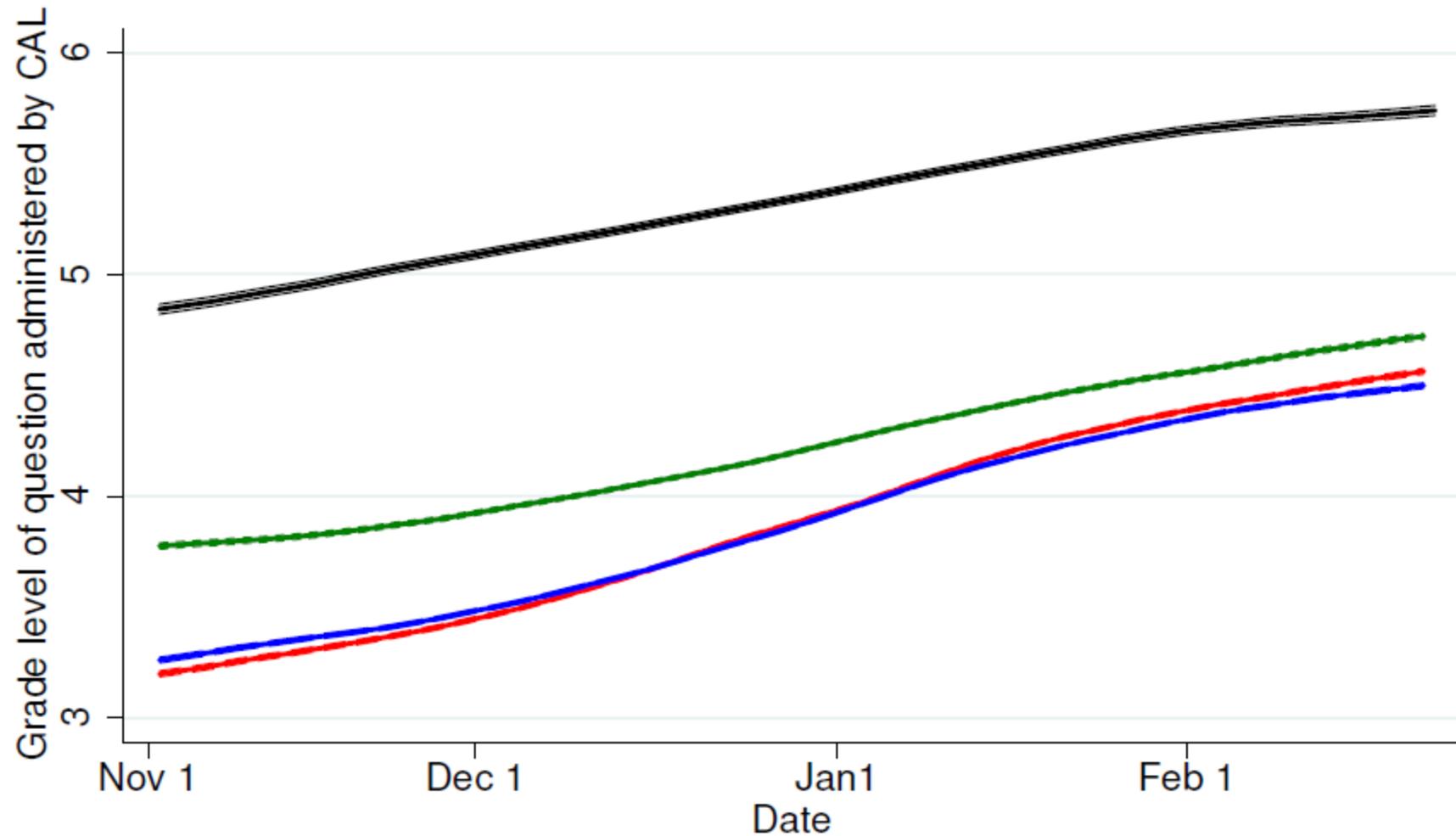
CAL caters to wide range of ability in a single session



This figure shows, for treatment group, the grade level of questions administered by the computer adaptive system to students **in a single day** (3 Nov 2015). The CAL system (a) allows for precise targeting to individual ability levels: (b) can cope with wide variation in

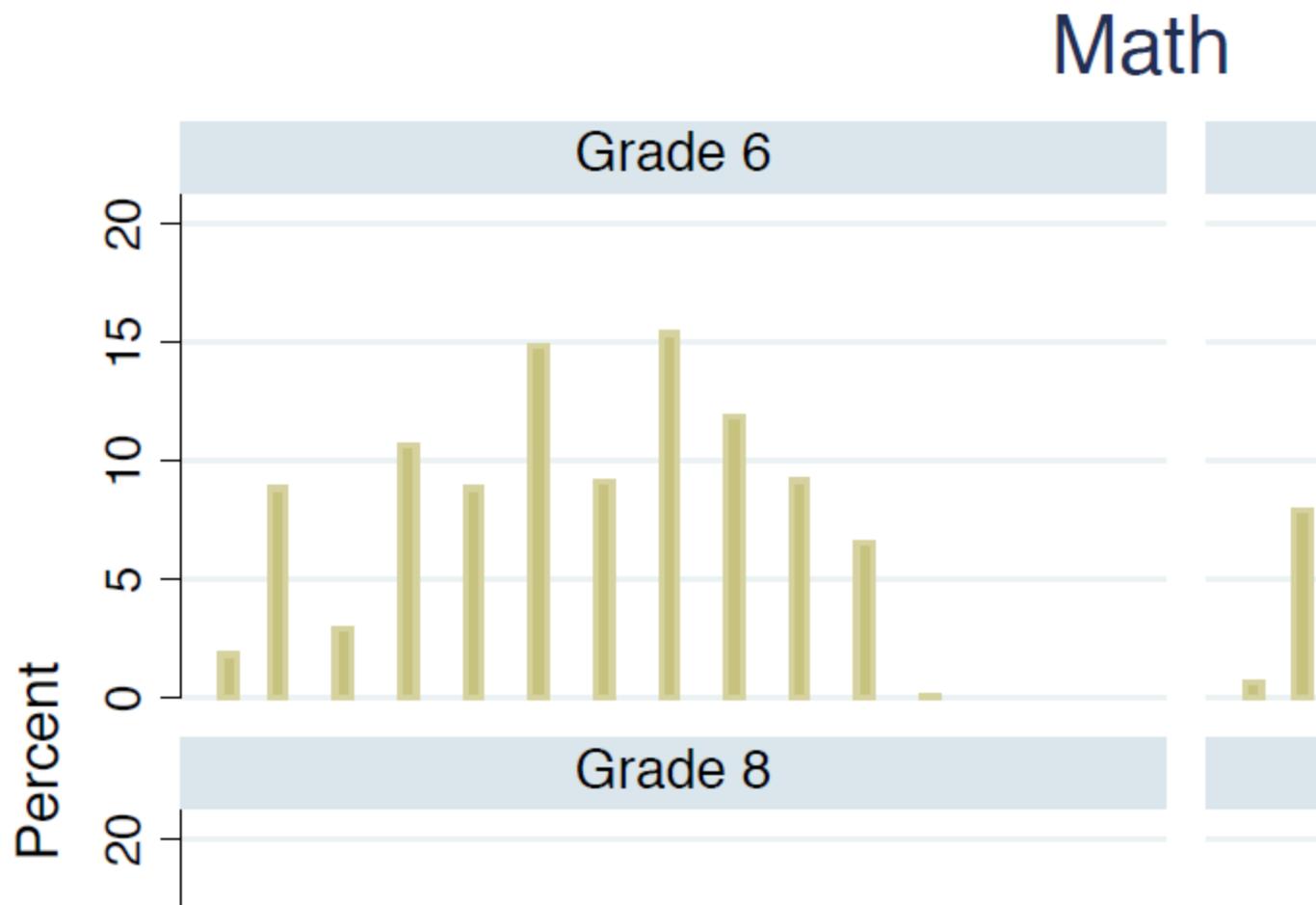
Students in all grades learn over the study period

The increase in learning is continuous *and continuously adapted to at individual level*



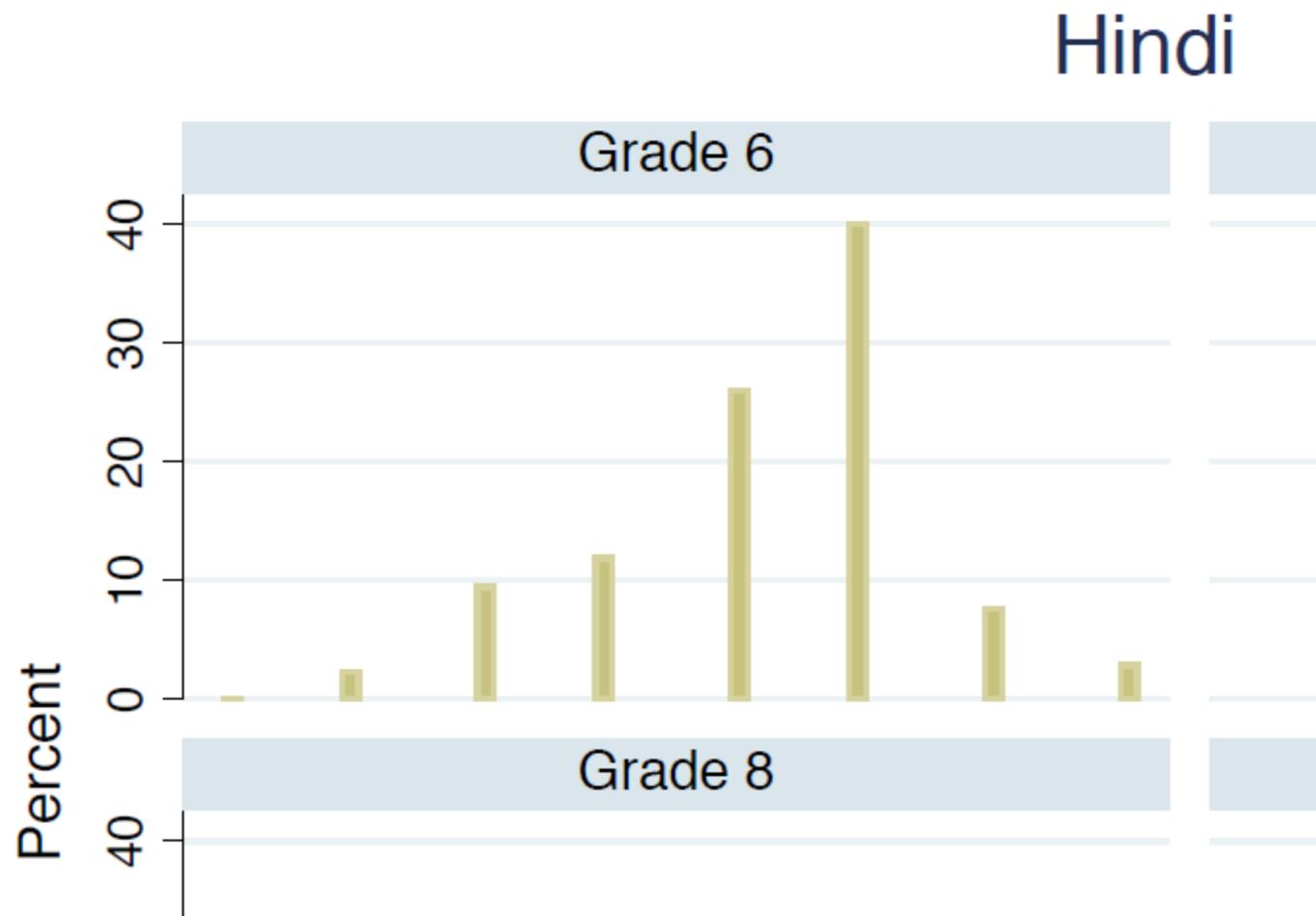
What was the Mindspark CAL sys

In math, very few questions at grade level



What was the Mindspark CAL sys

In Hindi, many more questions at grade level



Treatment effect on items linked to

	(1)	(2)
	<i>Dep var: Proportion of</i>	
	<u>Math</u>	
VARIABLES	At or above grade level	Below grade level
Treatment	0.0023 (0.039)	0.082*** (0.012)
Baseline math score	0.044 (0.025)	0.095*** (0.0056)
Baseline Hindi score		

Treatment effect on school exams

Across all subjects

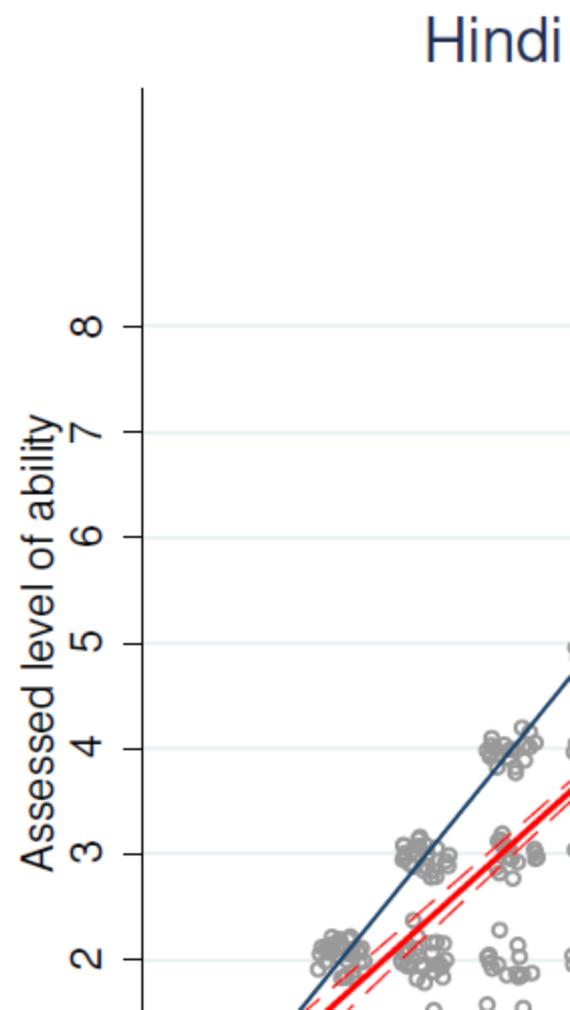
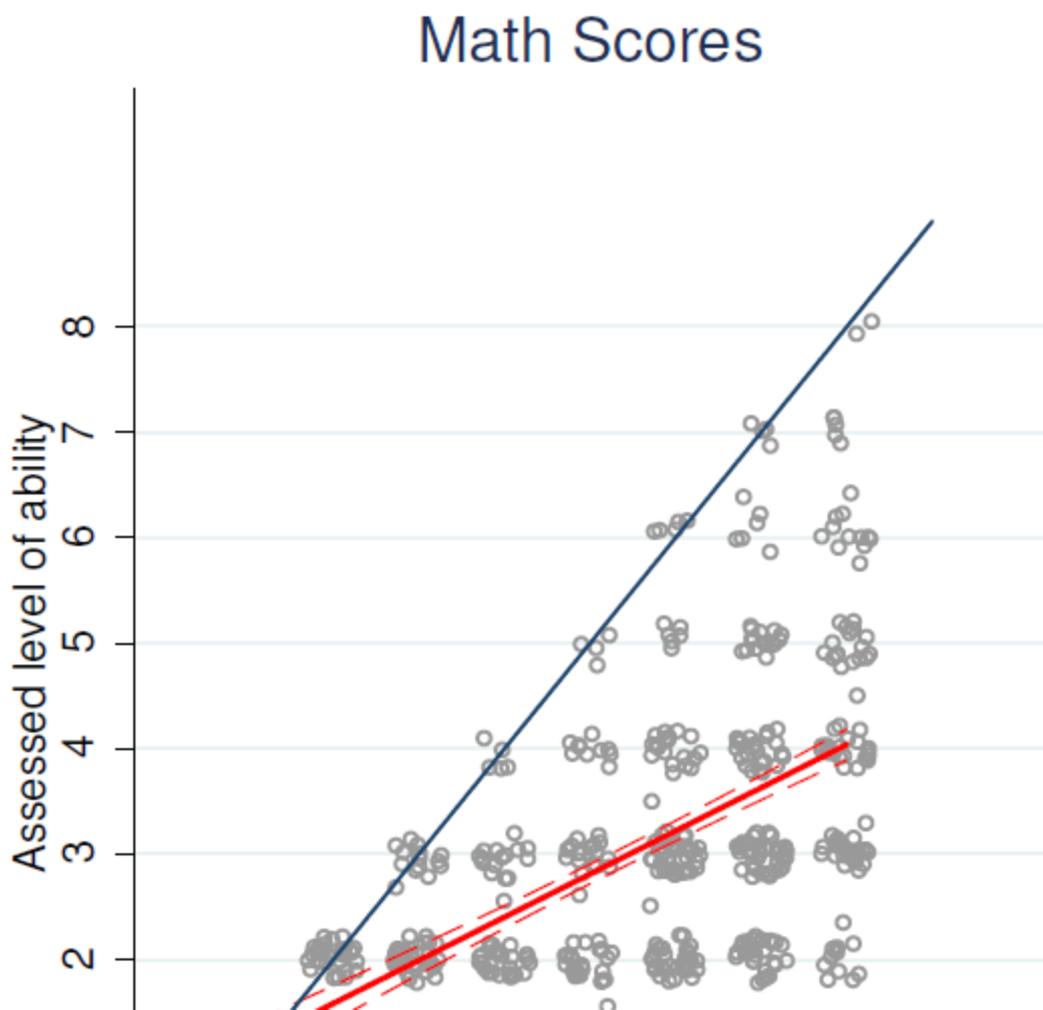
	(1)	(2)	(3)
VARIABLES	Hindi	Math	<i>Dep var: Star</i> Science
Treatment	0.19** (0.089)	0.058 (0.076)	0.077 (0.092)
Baseline Hindi score	0.48*** (0.094)		0.28*** (0.064)
Baseline math score		0.29*** (0.039)	0.10** (0.036)
Constant	0.40 (1.01)	0.14 (0.50)	0.88** (0.30)

Process Discovery for Scale

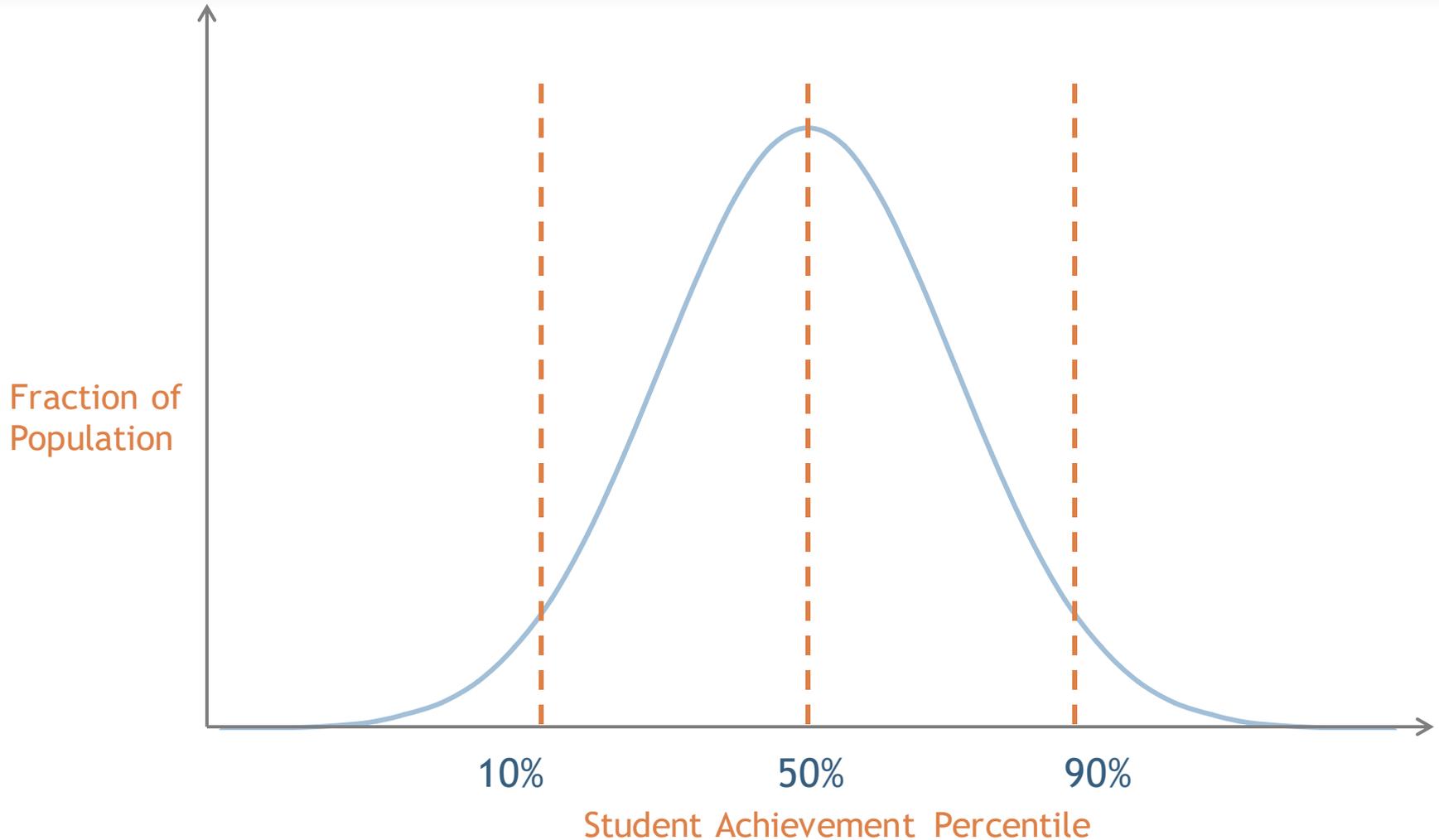
- The Delhi results are promising, but are best considered a “Proof of Concept” that large gains are possible in rapid time frames with a combination of the benefits of CAL and TaRL
- Can we replicate these results in government schools?
 - In Delhi, EI ran all the logistics and the after-school centers
 - Self-selected sample of interested students
 - Supplementing as opposed to substituting (cf with Berry & Mukherjee)
- Currently wrapping up a 3-year study of scale up reaching over 5000 students in 40 schools in state of Rajasthan
 - In-school model with Mindspark periods integrated into the timetable

Mismatch between grade levels and actual a

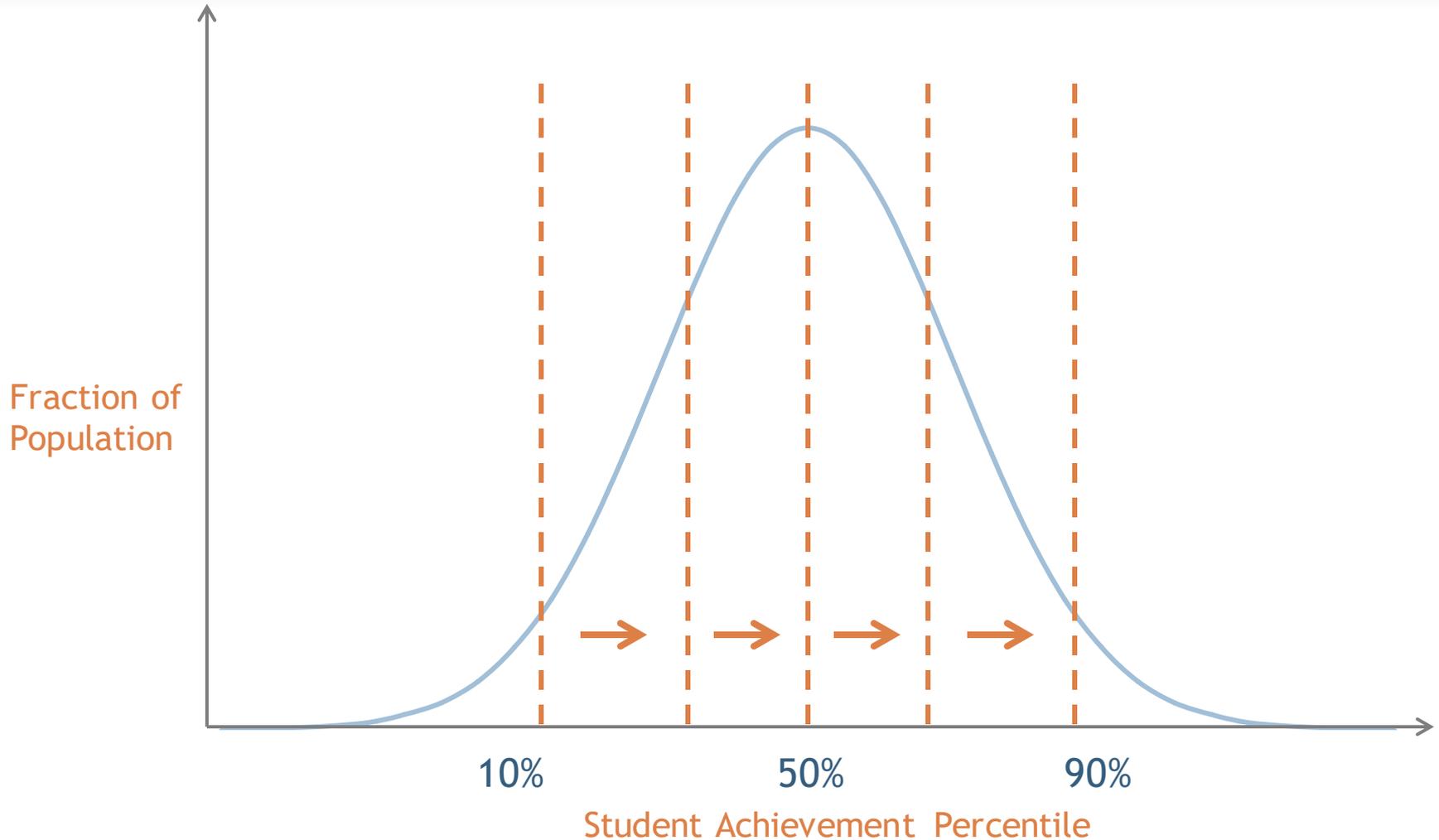
Learning deficits and within-grade dispersion in achievement



Selection Versus Developmental Paradigm in Education



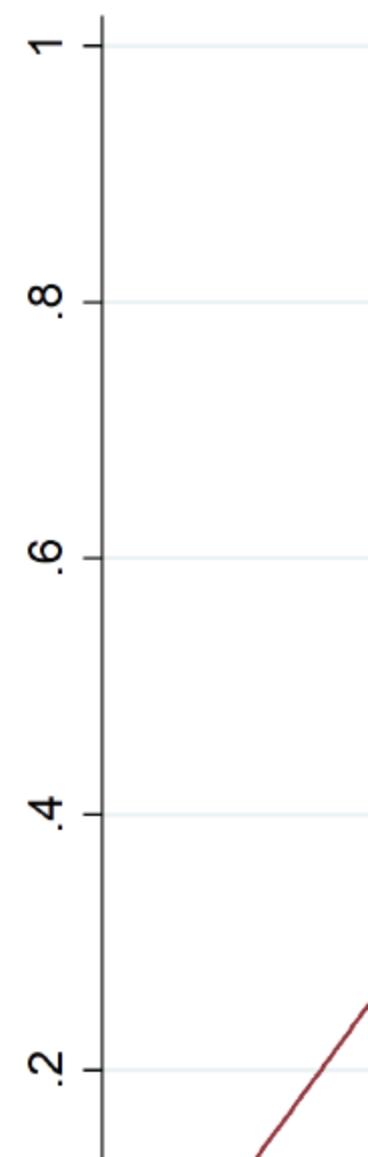
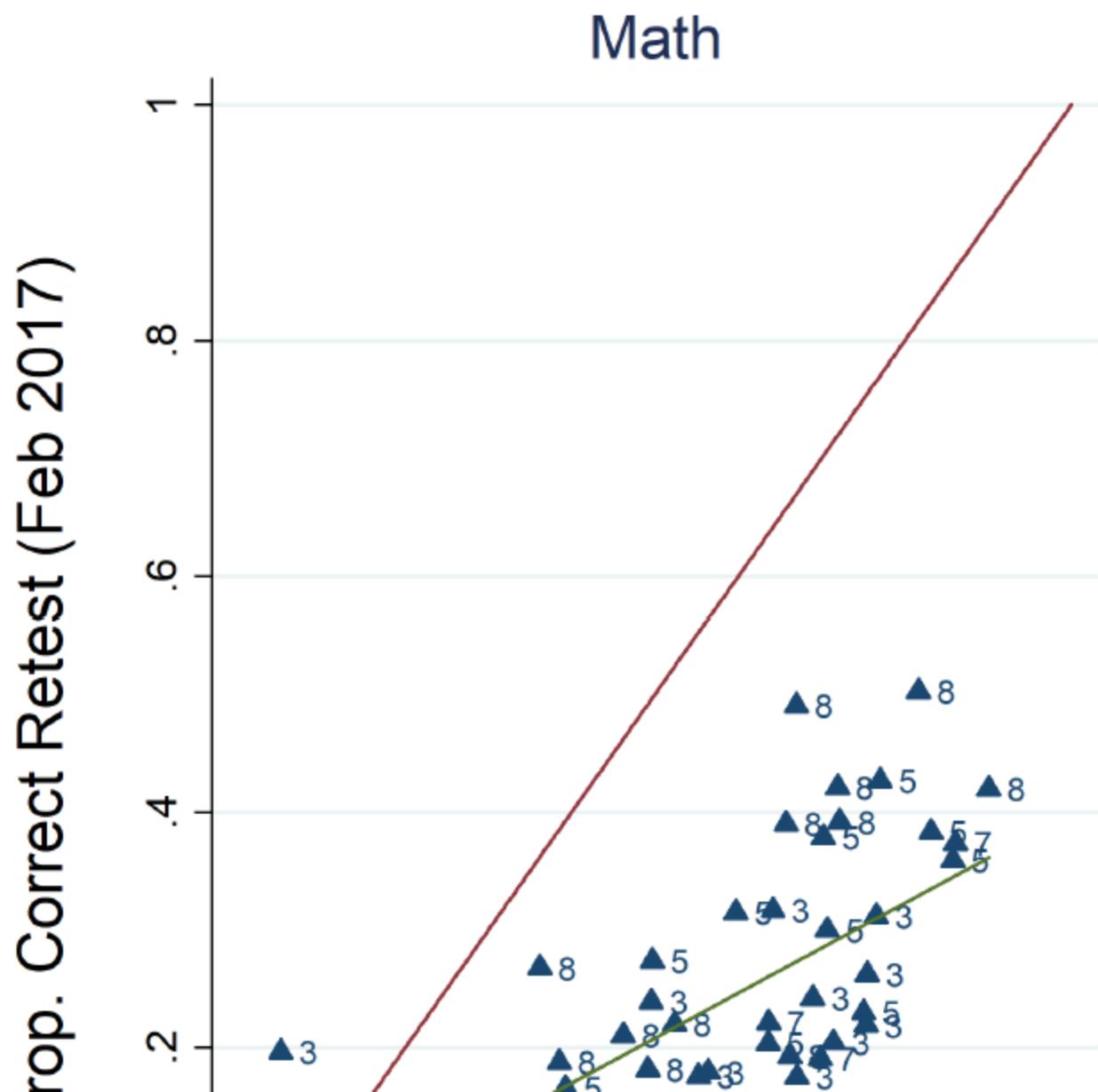
Moving from Selection to Human Capital at all Parts of the Distribution



A few recent papers of note

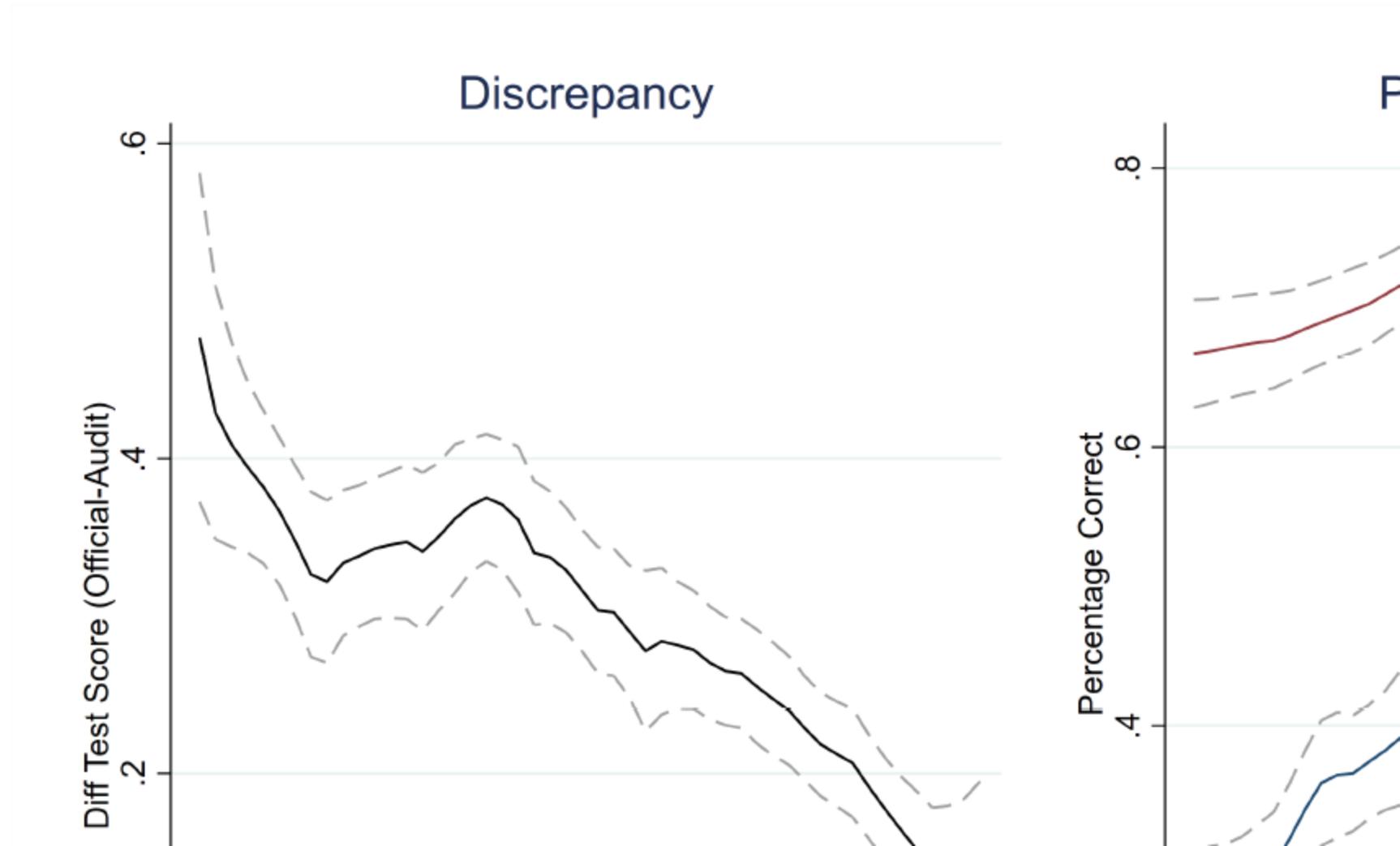
- Bianchi et al (2021)
 - Studies impact of connecting top teachers in China to rural students through broadband internet
 - Finds strong positive impacts on academic achievement; labor market outcomes
 - Finds these effects 7-10 years after the program suggesting long-term impacts
- Navarro-Sola (2021)
 - Studies impact of expanding junior secondary education in Mexico using *telesecundaria* – schools using televised lessons
 - Finds that each year of education increased earnings by 12.5-13.9%
- Beg et al (2021)
 - RCT in Pakistan found that adding expert-led curriculum-based videos to the class raised test scores, but that just giving the same content to students reduced test scores – suggesting key mediating role for teacher
- Derksen et al (2021)
 - Studies impact of Wikipedia access on learning
- Ferman et al (2021)
 - RCT of AI based feedback to students on writing – large positive effects

PP Jan 2017 vs. independent assessment in Feb 2017 (Math)



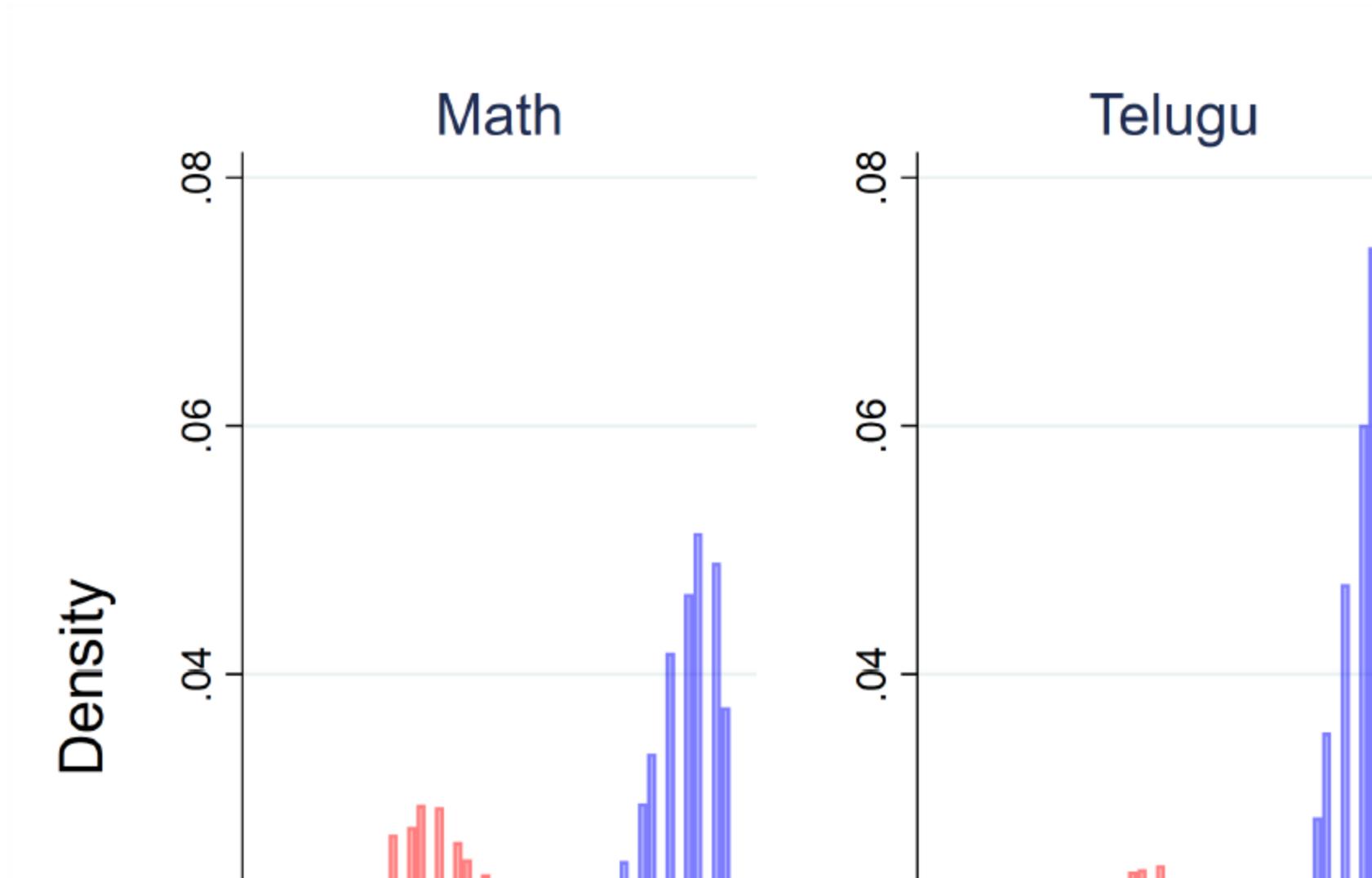
Overstated at all levels (but more for

Discrepancy between official and audit assessments



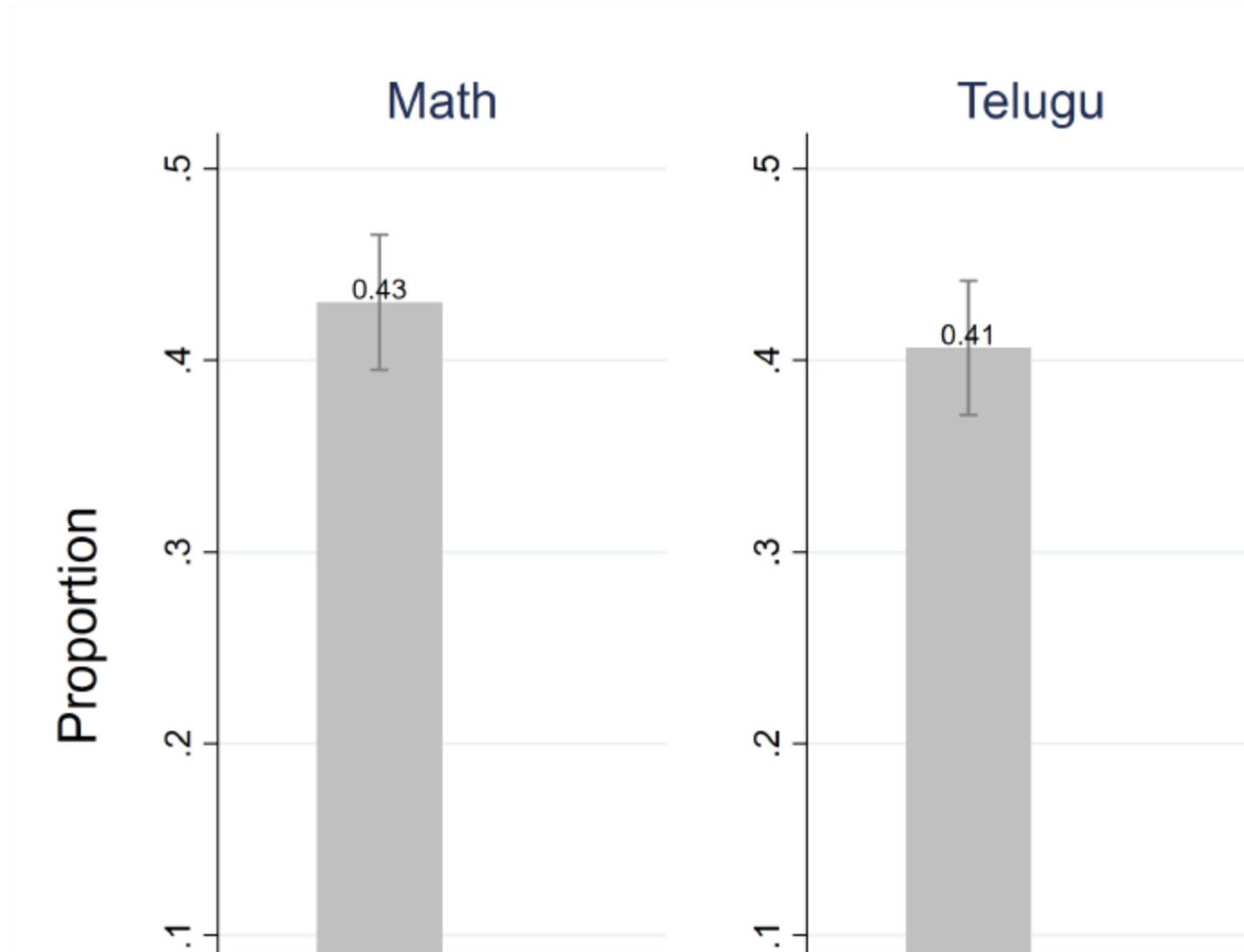
Results

Test score distribution in tablet and paper tests – Student



The main result

Paper-based tests much more likely to suggest cheating



Wrap up

- Technology has large potential to improve both pedagogy and governance in school systems
- But requires careful attention to what the binding constraints are and to using technology to alleviate these constraints
- Default focus in most governments is on hardware procurement (and putting pictures of politicians on laptops that are distributed!)
- This is unlikely to have much impact (OLPC, RJ examples)
- Pandemic is likely to have exacerbated inequalities (sometimes low-tech is best)
- Research frontiers include improvements in measurement, sophistication of what the computer does, engaging parents, reorienting teacher training, understanding and increasing student engagement, composite interventions that leverage and test complementarities