Comments at RISE Meeting

Karthik Muralidharan

UC San Diego, NBER, and J-PAL

18 June 2015

Explosion in Quantitative Research in Education in LICs in past decade

- One way of seeing is that there have been 6-10 review papers of the research on education in developing countries in just the *last two years*:
    - Muralidharan (2013) – Focused on India
    - Glewwe et al (2013) – School Inputs
    - Kremer et al (2013) – Short review (*Science*)
    - Krishnaratne et al. (2013) – 3ie review
    - Conn (2014) – Focused on sub-Saharan Africa
    - McEwan (2014) – Uses only RCT's
    - Murnane & Ganimian (2014) – NBER WP
    - Snilstveit et al. (2014) – Campbell Review (23 page proposal)
    - Glewwe & Muralidharan (2015)
    - Evans & Popova (2015) – includes a review of reviews!!

- Some takeaways:
    - Synthesizing research in 'meta analysis' is not easy
    - Even high-quality studies are very difficult to compare
    - Variation within 'theme' is often bigger than variation across 'themes'
    - But some broad themes do emerge

- Demand-side interventions
  - Conditional Transfers
  - Information to parents/communities
  - Student incentives
- School/student inputs
  - Buildings/Access
  - School grants, books, materials
  - Teachers (pupil teacher ratio, salary, training)
  - Deworming, school feeding
- Pedagogy
  - Teaching at the right level
  - Computers/technology
- Governance
  - Performance-linked pay
  - Contract teachers
  - School and village management committees
  - Choice and competition

Limitations: Interpreting Zero Effects

- In theory, this should just mean that the marginal product is zero
  - In practice, many different possibilities with different policy implications
- Four different studies in four different contexts all find close to zero impact of providing books & materials to students
  - But they point to four different reasons for non-impact!
- Sabarwal et al (2014) in Sierra Leone
  - Textbooks did reach the schools but were put in storage and not given to the kids! [Form of non-implementation]
- Das et al (2013) in India
  - Positive effect of books/materials in Year 1 of experiment, zero in Year 2
  - Households sharply reduced their own spending in Year 2 [Substitution]
- Glewwe et al (2009) in Kenya
  - No mean impact of free textbooks, but positive for top 20% of BL scorers
  - Did not alleviate binding constraints (inability to read) for most students
- Mbiti & Muralidharan (2015) in Tanzania
  - Zero effects of school grants (mostly spent on books and materials)
  - Strong positive interaction effect with teacher performance pay (significantly positive over and above the impact of performance pay alone)

Limitations: External Validity (both within & across contexts)

- "Treatment effect" is "Treatment" * "Context-specific unobervables"
  - No reason for these unobservables to be the same across contexts

- Many challenges to external validity even within the same context

- Representativeness of study universe

- Implementation quality (NGO vs. Government)

- Tweaking the policy (value of incentive, CCT, etc.)

- GE and political economy concerns with scale up (contract teachers)

- All these problems are magnified with external validity across contexts
  - Need to study multiple interventions in the same setting (Kenya, AP), and
  - Study the same intervention in multiple locations (seems to be less incentive compatible)

Implications for the RISE program

- Research is difficult
  - Let's not forget how we got to RCT world

- Economics profession broadly agrees that RCT's provide the best answer on the impact of a specific intervention in the specific setting

- But severe challenges to learning from this literature in a systematic way

- More work needs to be done to make them **more** useful

- Some key areas where RISE can help:
  - Create public goods that enable comparisons across studies – especially measurement tools and common scales for test scores, and processes
  - Create administrative public goods including longitudinal data (at least for a sample)
  - Support structures that combine iterative intervention design (by education experts), with rigorous evaluation (by evaluation experts), and to embed these findings in organizations to deliver these improvements at scale

Improving School Governance at Scale: Evaluating a System-level Reform

- Context is the Indian state of Madhya Pradesh (MP)
  - 5th largest state in India by population (75 million in 2011)
  - >120,000 government schools
  - One of the poorer states, both in incomes and in learning
  - Most populous tribal state
  - ASER data indicate that decline in learning levels among the steepest in the country

- Recent impetus towards government reforms in education
  - Pratibha Parv
  - Board exams in Grades 5 and 8
  - Work on improving school governance and quality
 - Led to the MP School Quality Assessment (QA) Program

- Developed in response to request from GoMP; with DFID support; led by ARK; inputs from OFSTED

- Establish a school governance architecture that provides
  - Quarterly Monitoring of government schools at all levels
  - Detailed School Quality Assessments that score schools on over 20 quality metrics in 7 broad categories; Work with stakeholders to build a school improvement plan (SIP)
  - Conduct quarterly follow-up visits to enable assessment of progress on SIPs
  - Monitoring of all assessment and feedback using dedicated website plus Android app.

- Builds on existing structures of state administration: CRCs, BRCs,
  - Helps in buy-in, scale-up, and policy relevance
 - Designed to scale from the outset with leadership FROM the government

- We worked with ARK and GoMP and convinced them to randomize the roll out across the entire state!
  - ~100 schools in 2013-14 (prototyping the program)
  - ~2000 schools in 2014 – 15 (successfully randomized)
  - ~20,000 schools in 2015 – 16 (randomized, roll out starts next month)
  - All schools in the state (~120,000) to be covered by 2018-19

- Not just randomizing lots of schools, but randomizing increasingly larger administrative units
  - Randomized at the cluster level (~40 schools each) in Phase 1
  - Randomized at the block (~400 schools each) in Phase 2
  - Allows us to progress from 'efficacy' to 'effectiveness' trials before scale up

- Detailed data collection on
  - Implementation quality (through Android MIS)
  - School processes; student test scores

Phase 1 (2014 – onwards)
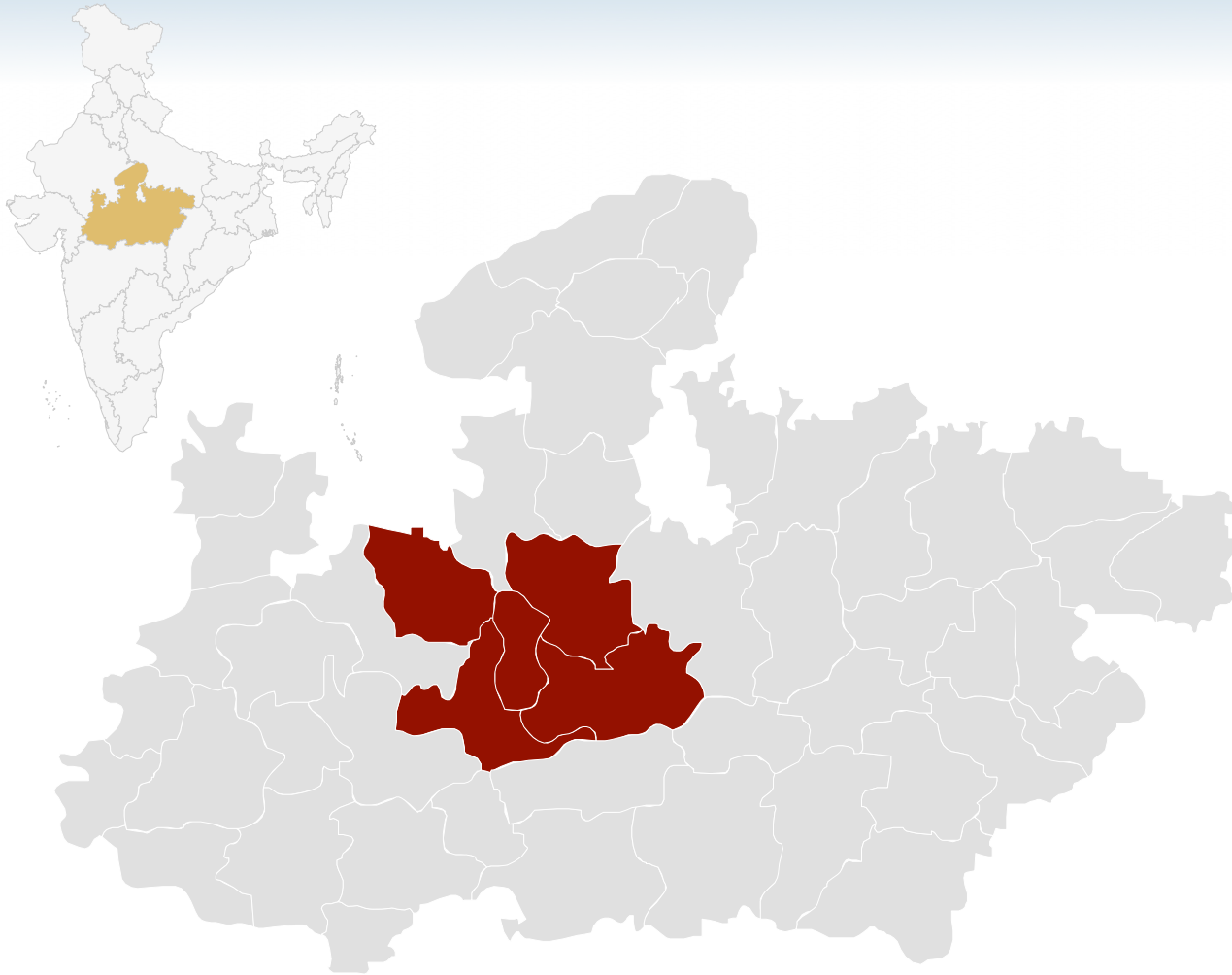
- Program rollout in five districts (Bhopal region)
- 2000 schools selected randomly (out of ~12000)
- Randomization at **academic cluster level**
- Balance on key infrastructure, inputs and full distribution of test scores
- Done with admin data – no expensive baseline needed
- Program spread across primary, middle and secondary schools
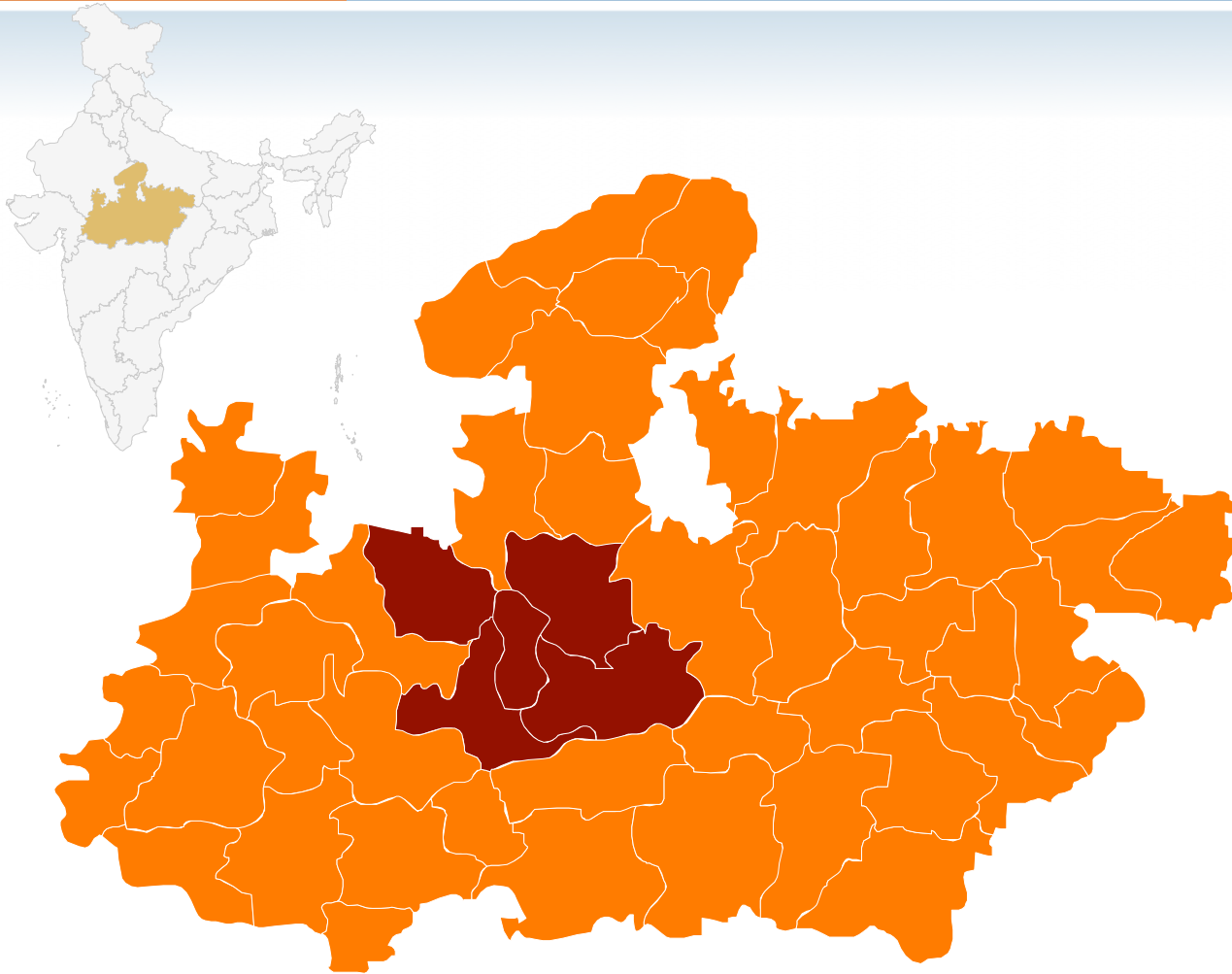
- Program rollout in all other districts (46 districts)
- 20k schools selected randomly
- Randomization at **block level**
- Balance on key infrastructure, inputs and full distribution of prior test scores (admin data)
- Program spread across primary, middle and secondary schools

Great Setting for "Systems Research"

- Results are informative and relevant regardless of whether positive or negative

- Positive results can accelerate scale up; non-results will take us to the process data to better understand why?

- Intervention is a composite (and differentiated!) one by construction
  - Focus less on a single "intervention" than on a *kaizen*-like management process of continuous improvement
  - Results can be interpreted as impacts of a "management" intervention in the public sector
  - Different binding constraints across settings (Rodrik); empower schools to focus on their most limiting weakness and work on those
  - Similar to Bloom et al (2013) with private management consulting

- First step in deep engagement with GoMP over a long time
  - Other topic areas of focus likely include ECE; PPP; Teacher selection, training, and performance management
  - Plan to embed qualitative researchers from the outset