

# **Learning in India: What Input and Outcome Data Can Tell Us**

**Abhijeet Singh**

Stockholm School of Economics

# Overview

- This talk is about **data** on the Indian **school system**
  - Won't be talking today about higher education
- Three main dimensions:
  - **Access:** who goes to school and where?
  - **Inputs:** What can we say about staffing, infrastructure, spending etc.?
  - **Outcomes:** What skills do children have?
- Specifically, on the following questions:
  - What data exist?
  - Can they be accessed?
  - Can they be trusted?
  - Where can we go from here in terms of data availability?

# Intended audience

- Anyone who intends to work with microdata directly
  - Not just look at published summary statistics but go further in validating the data, looking at patterns across variables etc.
- Typical use cases: data journalism, policy inputs, student theses
- **NOT** primarily a focus on how to do frontier research in this field
  - That has very separate incentives, requirements
- Focus on understanding broad trends and cross-sectional relationships

Unified Digital Information on  
School Education  
(UDISE+)

# Dataset 1: UDISE+

- Starting point for understanding the school system is U-DISE+
  - Unified Digital Information on School Education ([udiseplus.gov.in](http://udiseplus.gov.in))
- Register of all recognized schools across all types of management
  - Also a category for unrecognized schools but these are few in the data
  - This used to be a much bigger issue in the past but is less of a constraint now
- What information does this have?
  - Enrollment (by grade, sex, social category), staffing, infrastructure
  - School establishment details
- What does this allow us to say?
  - 265 million children in K-12, 88 million in private schools
  - Core tabulations provided in annual reports (on Ministry of Education website)

Unified District Information System for Education Plus  
(UDISE+)

**Academic Year 2021-22**

*Reference Date: 30th September 2021*

**For Higher Secondary Schools of Grade I to XII**

Department of School Education and Literacy  
Ministry of Education  
**Government of India**

Sections	Details
Section 1	School Profile (Location, Management, Medium of Instruction etc)
Section 2	Physical Facilities, Equipment, Computer and Digital initiatives
Section 3	Teaching and Non-Teaching Staff
Section 4	Enrolment and Repeaters
Section 5	Incentives and Facilities Provided (Only for Govt. and Govt. Aided Schools)
Section 6	Annual Examination Results
Section 7	Board Examination Result
Section 8	Receipts and Expenditure
Section 9	Vocational Education under NSQF at Institutional level
Section 10	PGI and Other Indicators (Only For Government and Government Aided Schools)
Section 11	School Safety
Section 12	Gifted Children

# Pros and cons

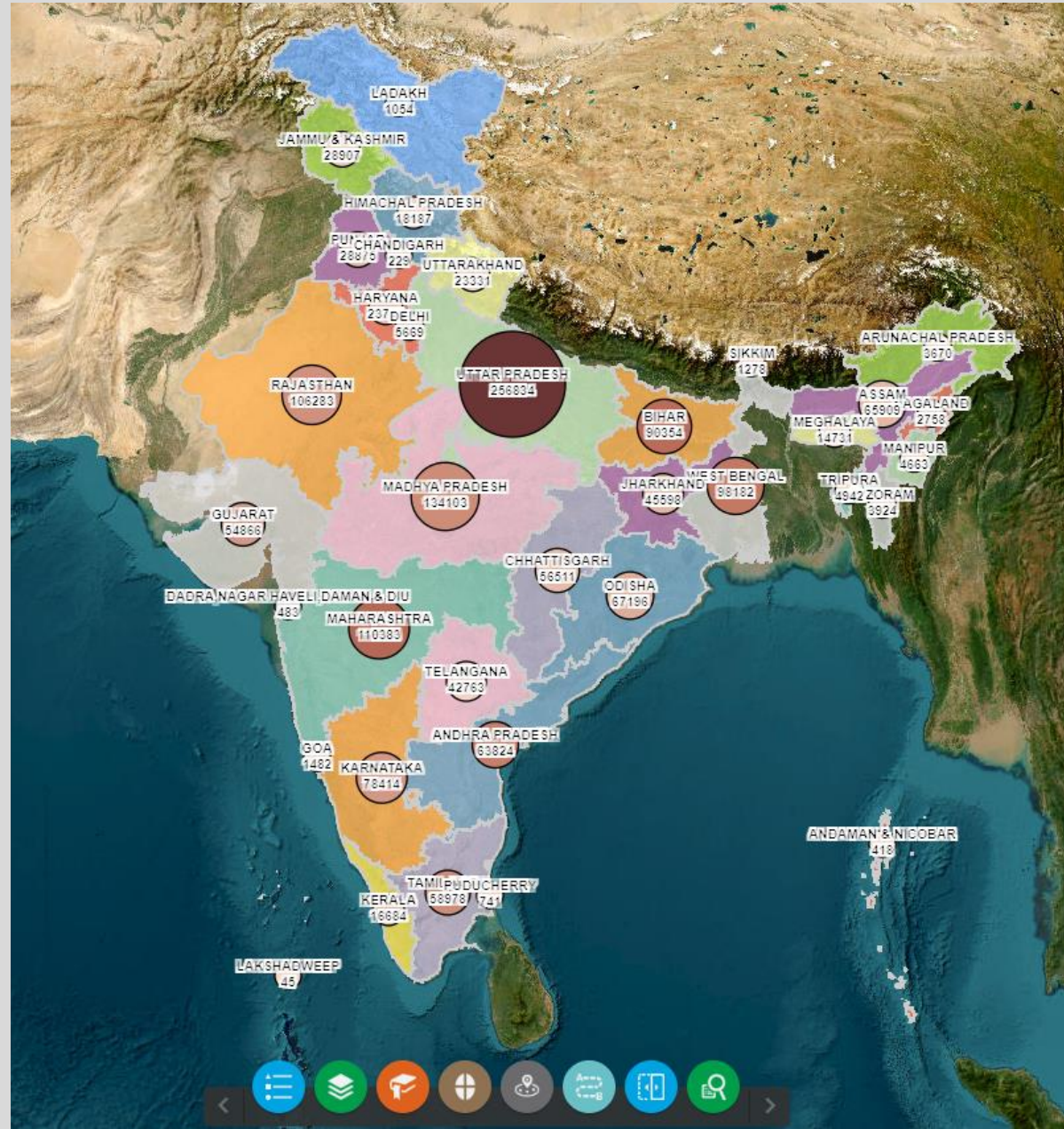
## **Pros:**

- U-DISE is particularly useful because the micro-data are available
- Covers the universe of schools
- Available over a long time frame
- You can build a panel at school level (or at any level of geography)
- On core variables (enrollment, staffing, infrastructure), it's reasonably reliable

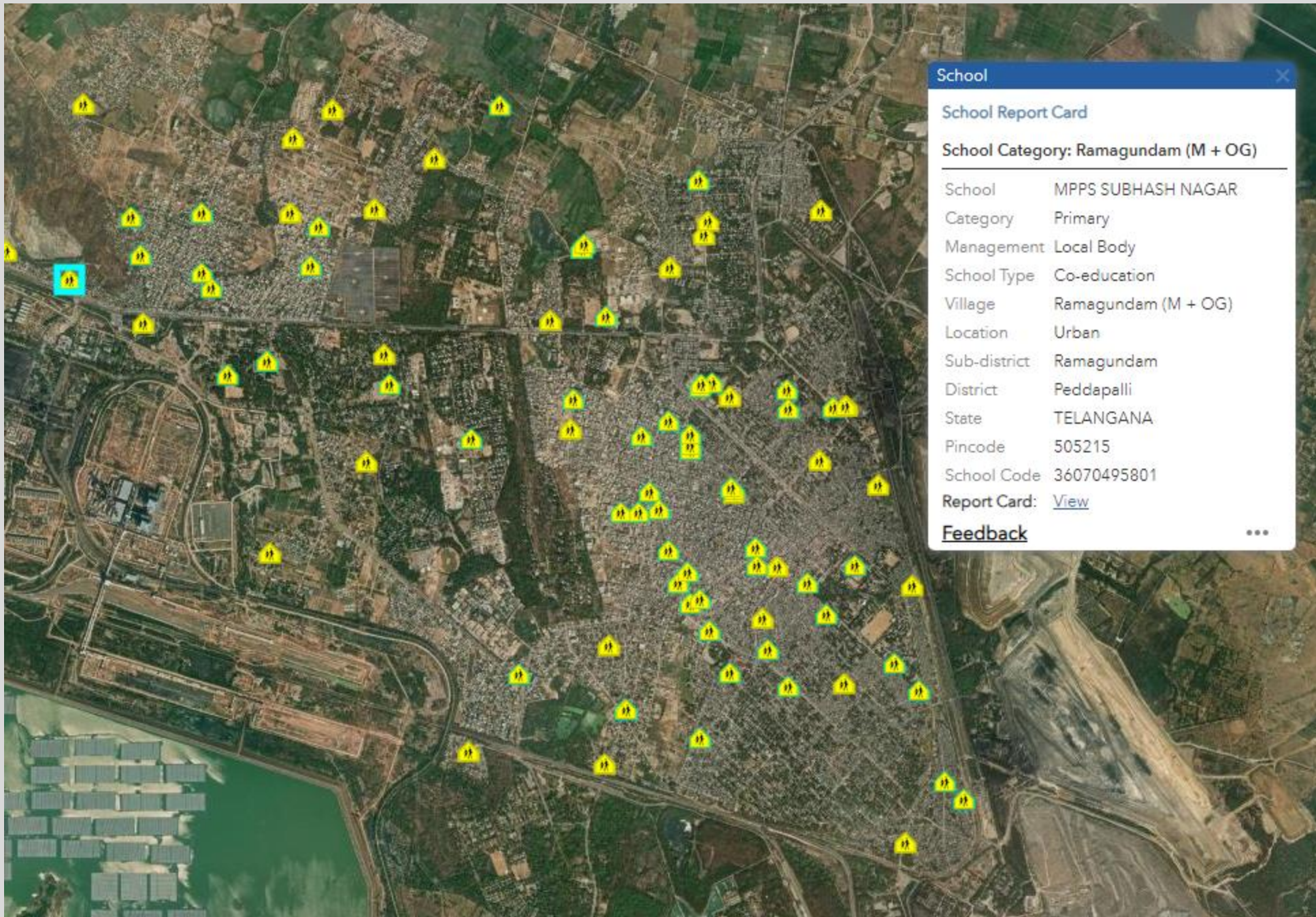
## **Cons:**

- No reliable data on learning outcomes
- Limited information on student characteristics
- No information on school fees in the publicly released data

# Spatial information linked to UDISE+ (schoolsGIS.nic.in)







School ✕

School Report Card

School Category: Ramagundam (M + OG)

School	MPPS SUBHASH NAGAR
Category	Primary
Management	Local Body
School Type	Co-education
Village	Ramagundam (M + OG)
Location	Urban
Sub-district	Ramagundam
District	Peddapalli
State	TELANGANA
Pincode	505215
School Code	36070495801

Report Card: [View](#)

Feedback ⋮

# What have people used it for? (Kingdon, 2020)

**Table 4. The emptying of government schools: Change over time in the number of 'small' and 'tiny' government schools**

Total number of pupils in the school as a whole:	Number of Schools	Percentage of total govt. schools	Number of Teachers	Total Enrolment	Average pupils per school	Pupil teacher ratio	Teacher Salary Expenditure (Rs. Crore)	Govt. Annual Per-pupil Salary Exp. (Rupees)	Govt. Monthly Per-pupil salary Exp. (Rupees)
2010-11									
Zero	4,435	0.43	14,304	0	0	0	503	-	-
5 or Less	8,675	0.84	21,277	15,333	1.8	0.7	748	4,88,101	40,675
10 or Less	21,008	2.03	42,843	1,18,166	5.6	2.8	1,507	1,27,530	10,628
20 or Less	71,189	6.87	1,38,033	9,20,254	12.9	6.7	4,855	52,760	4,397
50 or Less	3,13,169	30.24	6,33,323	95,10,902	30.4	15.0	22,277	23,422	1,952
All govt. schools	10,35,602	100.00							
2016-17									
Zero	6,714	0.64	11,791	0	0.0		677	-	-
5 or Less	14,991	1.44	26,043	29,638	2.0	1.1	1,495	5,04,256	42,021
10 or Less	36,365	3.48	68,586	2,04,421	5.6	3.0	3,936	1,92,539	16,045
20 or Less	1,16,307	11.14	2,38,213	14,65,423	12.6	6.2	13,670	93,285	7,774
50 or Less	4,17,193	40.00	9,85,051	1,18,59,775	28.4	12.0	56,529	47,664	3,972
All govt. schools	10,43,939	100.00							

Source: [www.statereportcards/rawdata/201011](http://www.statereportcards/rawdata/201011) Data analysed here are for 21 major states (counting Telengana as a separate state).

# Sanitation and Education<sup>†</sup>

By ANJALI ADUKIA\*

*I explore whether the absence of school sanitation infrastructure impedes educational attainment, particularly among pubescent-age girls, using a national Indian school latrine construction initiative and administrative school-level data. School latrine construction substantially increases enrollment of pubescent-age girls, though predominately when providing sex-specific latrines. Privacy and safety appear to matter sufficiently for pubescent-age girls that only sex-specific latrines reduce gender disparities. Any latrine substantially benefits younger girls and boys, who may be particularly vulnerable to sickness from uncontained waste. Academic test scores did not increase following latrine construction, however. Estimated increases in enrollment are similar across the substantial variation in Indian district characteristics. (JEL H75, H76, I21, I25, J16, O15, O53)*

Adukia, A. (2017). Sanitation and education. *American Economic Journal: Applied Economics*, 9(2), 23-59.

# Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.5m Neighborhoods\*

Anjali Adukia<sup>†</sup> Sam Asher<sup>‡</sup> Kritarth Jha<sup>§</sup> Paul Novosad<sup>¶</sup> Brandon Tan<sup>||</sup>

July 2022

## Abstract

There is little quantitative evidence on the role that neighborhood settlement patterns play in mediating inequality in rapidly urbanizing lower-income countries. This paper helps to close this gap by analyzing settlement patterns, access to public services, and economic outcomes across 1.5 million neighborhoods for two of India's marginalized communities: Scheduled Castes (SCs) and Muslims. Patterns of segregation and unequal access to public services in India's cities largely replicate the striking inequalities in its villages. Segregation of SCs and Muslims in both cities and rural areas is substantial in magnitude. We find that public schools and hospitals are systematically located *away* from neighborhoods where Muslims and SCs are concentrated. These disparities are driven almost entirely by differences across neighborhoods and within towns, and are thus hidden by the more aggregated data typically used to study inequality. Inequality in public service is thus driven by local decision-making — the most informal and least systematically observed level of government. Children from all groups who grow up in urban minority neighborhoods attain less schooling, even after controlling for parent education, and household and neighborhood consumption. Unequal access to public facilities in India's highly segregated neighborhoods may be a significant contributor to disadvantages faced by marginalized groups.

# What have I used it for?

## **Sampling frame**

- Even though most of my research uses primary data, you need an initial sample and randomization
- U-DISE is comprehensive for all government schools and most private schools

## **Basic information on school characteristics**

- Registered enrollment
- Number of teachers
- Number of rooms
- Highest grade
- Official medium of instruction

# Annual Survey of Education Reports (ASER)

# Annual Survey of Education Report

- The big black hole in Indian education data is student learning
  - There is no publicly available dataset that can tell us what the median school kid in India knows or can do!
- The closest are the data from ASER reports produced since 2005 by Pratham covering **rural** areas
- Volunteer-led survey covering rural areas of most districts in India
  - Two stage sample of villages
  - Sample of kids
  - Basic learning outcomes collected regardless of grade enrolled in
- **Single biggest public good to education research in India**

# Sample

- In 2022, covered 616 districts, 19k villages, 374k households, 700k children
- Sampling using Census 2011 sample frame
  - 30 villages per district
  - 20 households randomly selected per village
  - All children 3-16 y are surveyed
  - All children 5-16 y are assessed on basic skills
- Conducted by different nodal district organizations (DIETs, NGOs..)



# Quick glimpse - ASER tasks : Reading & Arithmetic

## READING TASKS

Std II level text

राजू नाम का एक लड़का था। उसकी एक बड़ी बहन व एक छोटा भाई था। उसका भाई गाँव के पास के विद्यालय में पढ़ने जाता। वह खूब मेहनत करता था। उसकी बहन बहुत अच्छी खिलाड़ी थी। उसे लंबी दौड़ लगाना अच्छा लगता था। वे तीनों रोज़ साथ-साथ मौज-मस्ती करते थे।

Std I level text

रानी नदी किनारे रहती है।  
नदी में बहुत मछलियाँ हैं।  
रानी उनको दाना देती है।  
वे सब मज़े से दाना खाती हैं।

Letters

म र ड  
ह च  
ल ब न  
क य

Words

गाना खुश  
मौसी  
पैर झोला  
किला  
आग मोर

## ARITHMETIC TASKS

Number Recognition/  
अंक पहचान 1-9

3

7

1

4

Number Recognition/  
संख्या पहचान 11-99

65

38

92

23

Subtraction/घटाव  
(2 digit with carry over)

52  
- 24

76  
- 47

Division/भाग  
(3 digit by 1 digit)

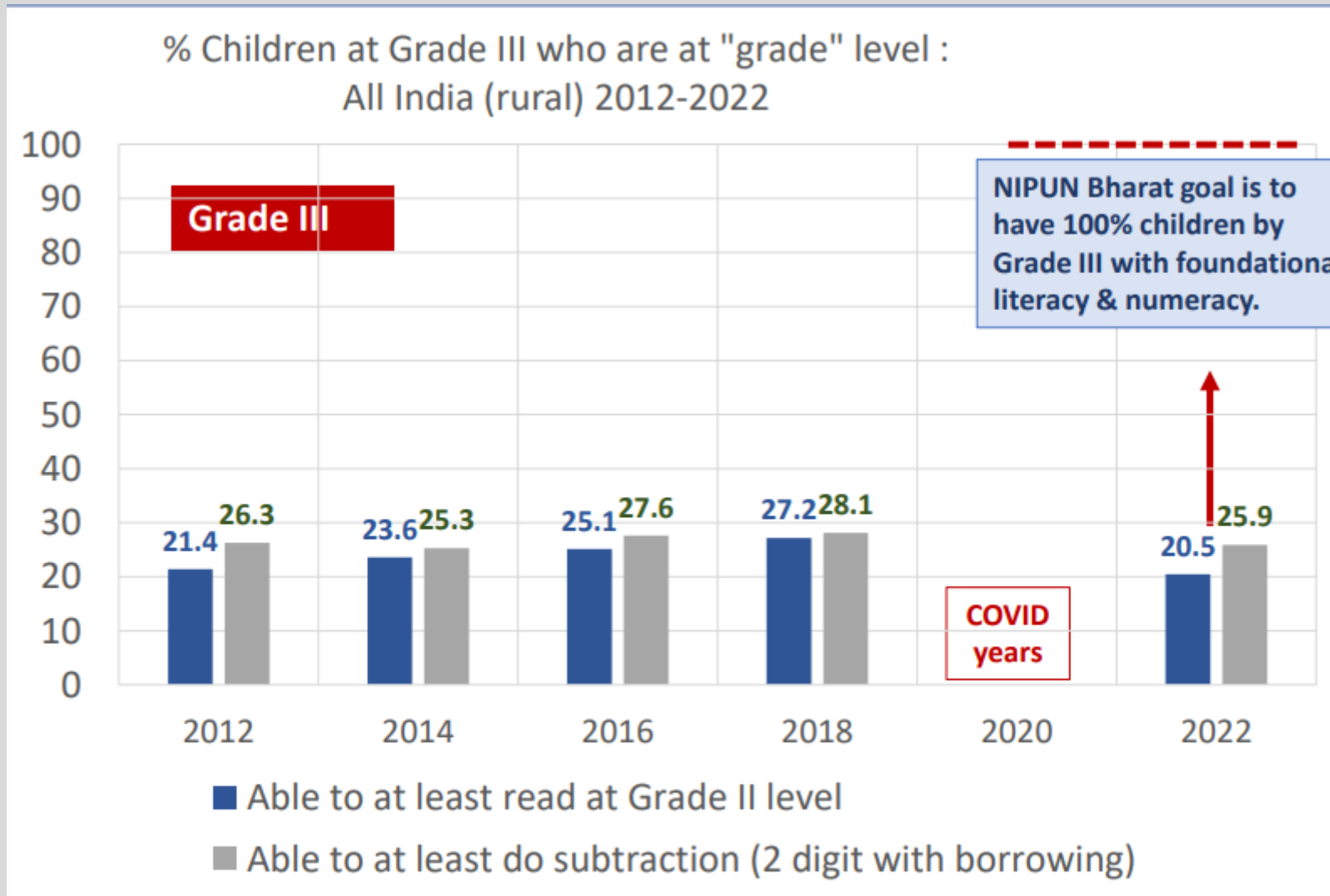
6 919

ASER is a household survey. Each child assessed one on one. S/he is marked at the highest level that s/he is able to do. Reading tasks are available in all regional languages.

# What have people used it for?

- The biggest (and most influential) use has been to show us descriptive statistics of what learning in Indian schools is like
  - Levels and trends of enrolment, attendance, school facilities and **learning**
- The “learning crisis” would be unknown in the absence of ASER
  - Single most important subject when thinking of school education in India
  - News typically focuses on Board Exams, IITs, IIMs
  - This is about foundational skills and has determined policy thinking over the long haul (in India and globally)
- It’s also been used in lots of research papers

# Big push needed to reach foundational literacy and numeracy goals



*By the end of Grade II in India, children are expected to be able to read a simple text fluently and also be able to do basic operations like subtraction. Hence, it is possible to use ASER data as a 'proxy' for the proportion of children who are at 'grade level' by the time they have reached the middle of the school year in Grade III.*



## School feeding and learning achievement: Evidence from India's midday meal program



Tanika Chakraborty<sup>a</sup>, Rajshri Jayaraman<sup>b,\*</sup>

<sup>a</sup> IIM Calcutta, India

<sup>b</sup> ESMT Berlin, Germany

### ARTICLE INFO

*JEL classification:*

I21

I25

O12

*Keywords:*

School feeding

Learning

Midday meal

Primary school education

### ABSTRACT

We study the effect of the world's largest school feeding program on children's learning outcomes. Staggered implementation across different states of a 2001 Indian Supreme Court Directive mandating the introduction of free school lunches in public primary schools generates plausibly exogenous variation in program exposure across different birth cohorts. We exploit this to estimate the effect of program exposure on math and reading test scores of primary school-aged children. We find that prolonged exposure to midday meals has a robust positive effect on learning achievement. We further investigate various channels that may account for this improvement including complementary schooling inputs, heterogeneous responses by socio-economic status, and intra-household redistribution.

# Data availability

- Microdata are available from the ASER Centre
  - Very willingly provided, decent documentation
  - Broad levels seem to agree with other independent data
  - Very usable datasets
- 
- Combined with substantial coverage and repeated measurement, this ends up being the main reference for education outcomes in India

# Issues

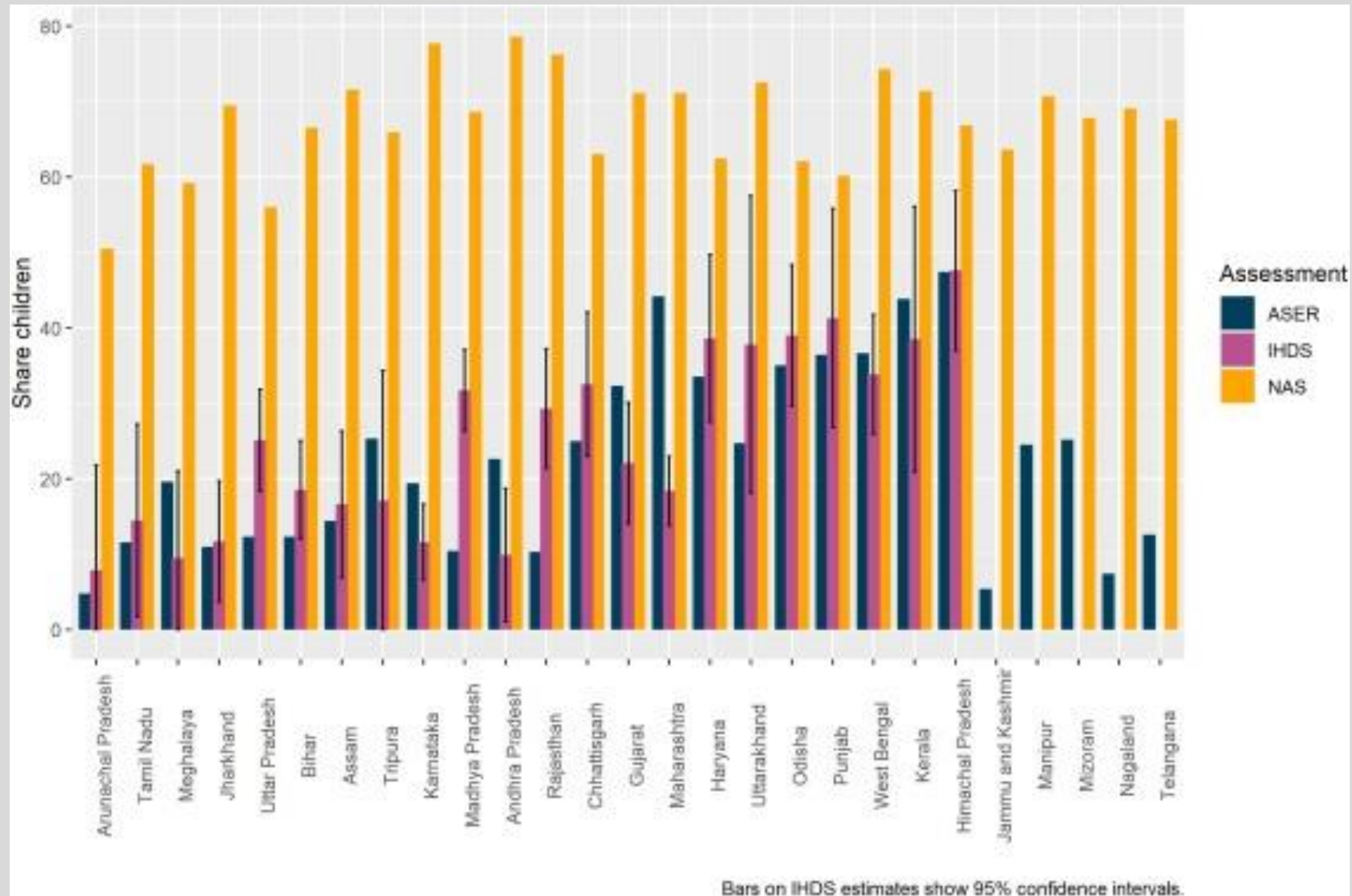
- It's a repeated cross-section which limits the analyses possible
  - Apparently there is a panel element to villages but village identifiers aren't available
  - You can study policies that vary across states or districts
  - Also can't be mapped on to GIS information
- The achievement tests are simple and offer only a few categories
  - This is great for the principal purpose of the survey
  - But good education research needs well-distributed scores
  - **Tip:** do not use z-scores (standard deviations) if you use ASER as an outcome, express effects in the competences directly
- Figures at district and state level are very “jumpy”, more than sampling variation can account for (*Johnson and Parrado, 2021*)
  - You need to think about measurement error seriously with panel data

# Administrative Data on Student Learning

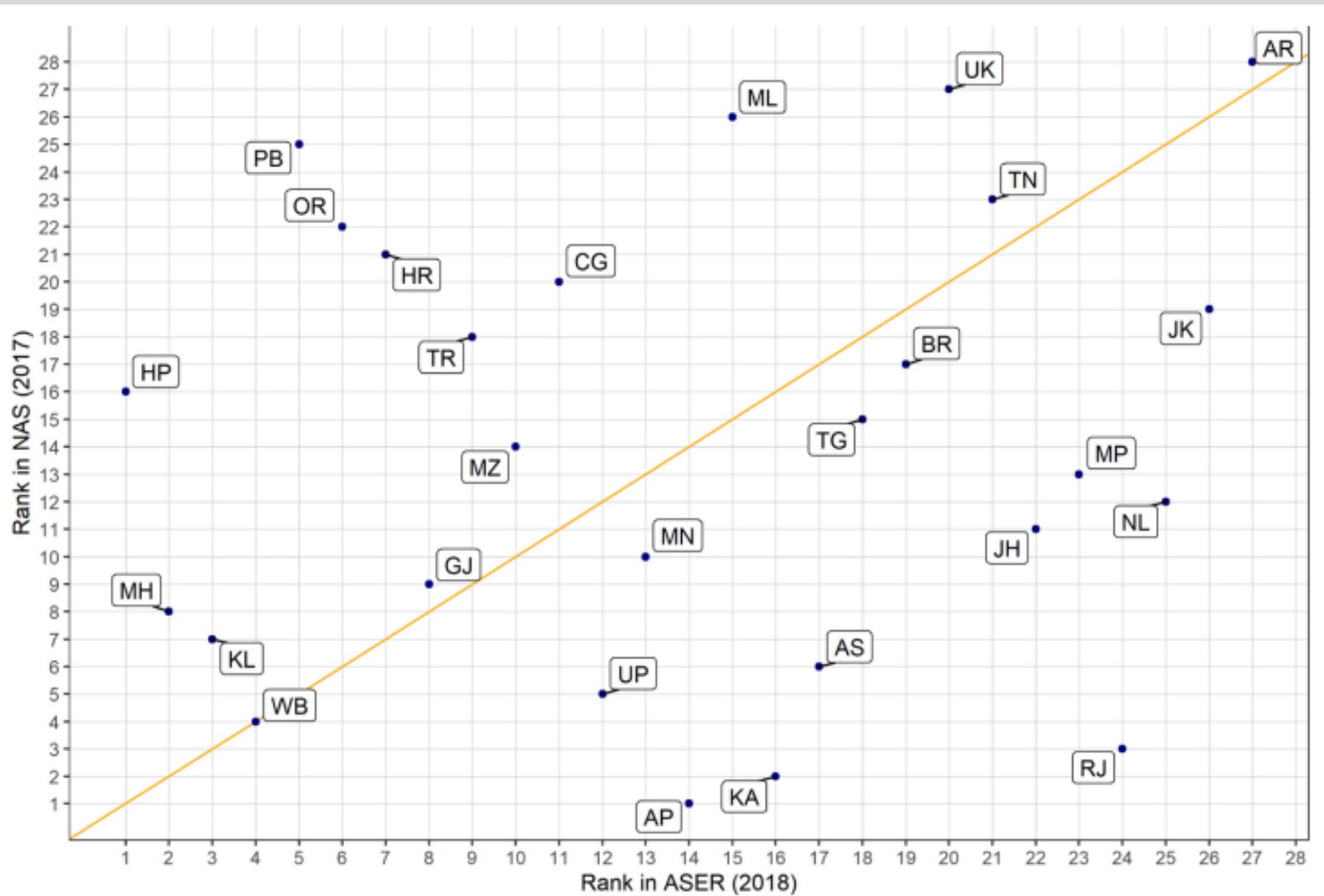
# What about the National Achievement Survey?

- The Indian state **also** thinks learning should be measured
- The National Achievement Survey is a nationally representative large-scale survey of students' learning
  - Intended to be premier source for school learning information in India
  - Tests students in Grades 3, 5, 8
  - Detailed sampling documentation, advanced psychometric analyses of tests
- Unfortunately, it seems to be completely unreliable and unverifiable
  - Micro data is not available
  - The patterns in NAS seem very counter-intuitive





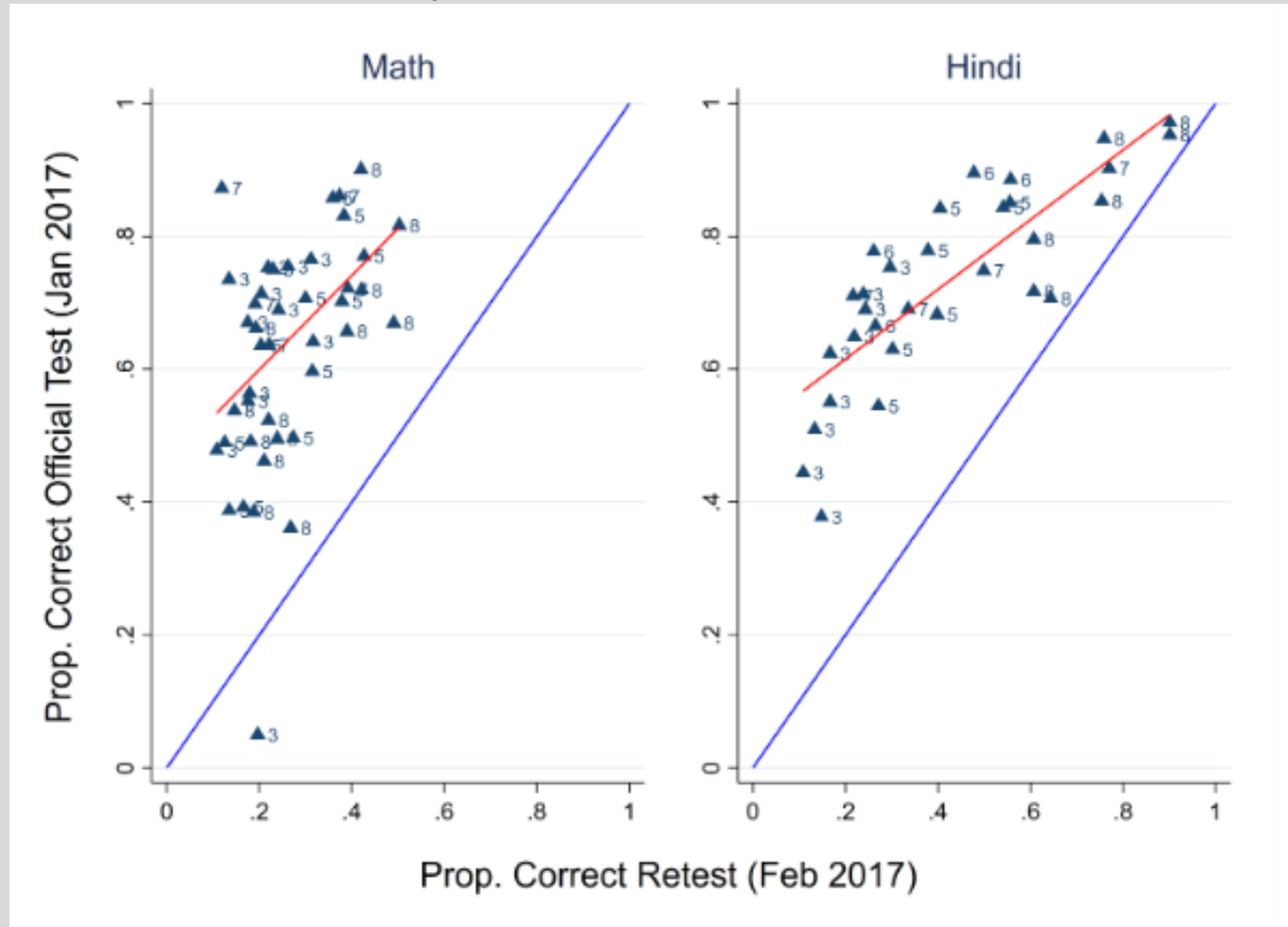
Johnson, D., & Parrado, A. (2021). Assessing the assessments: Taking stock of learning outcomes data in India. *International journal of educational development*, 84, 102409.



Johnson, D., & Parrado, A. (2021). Assessing the assessments: Taking stock of learning outcomes data in India. *International journal of educational development*, 84, 102409.

# State-led assessments are manipulated

## Pratibha Parv in Madhya Pradesh



Singh, A. (2020). Myths of official measurement: Auditing and improving administrative data in developing countries. Research on Improving Systems of Education (RISE) Working Paper, 42.

## **The tyranny of metrics: Why we learning nothing from the learning outcome data**

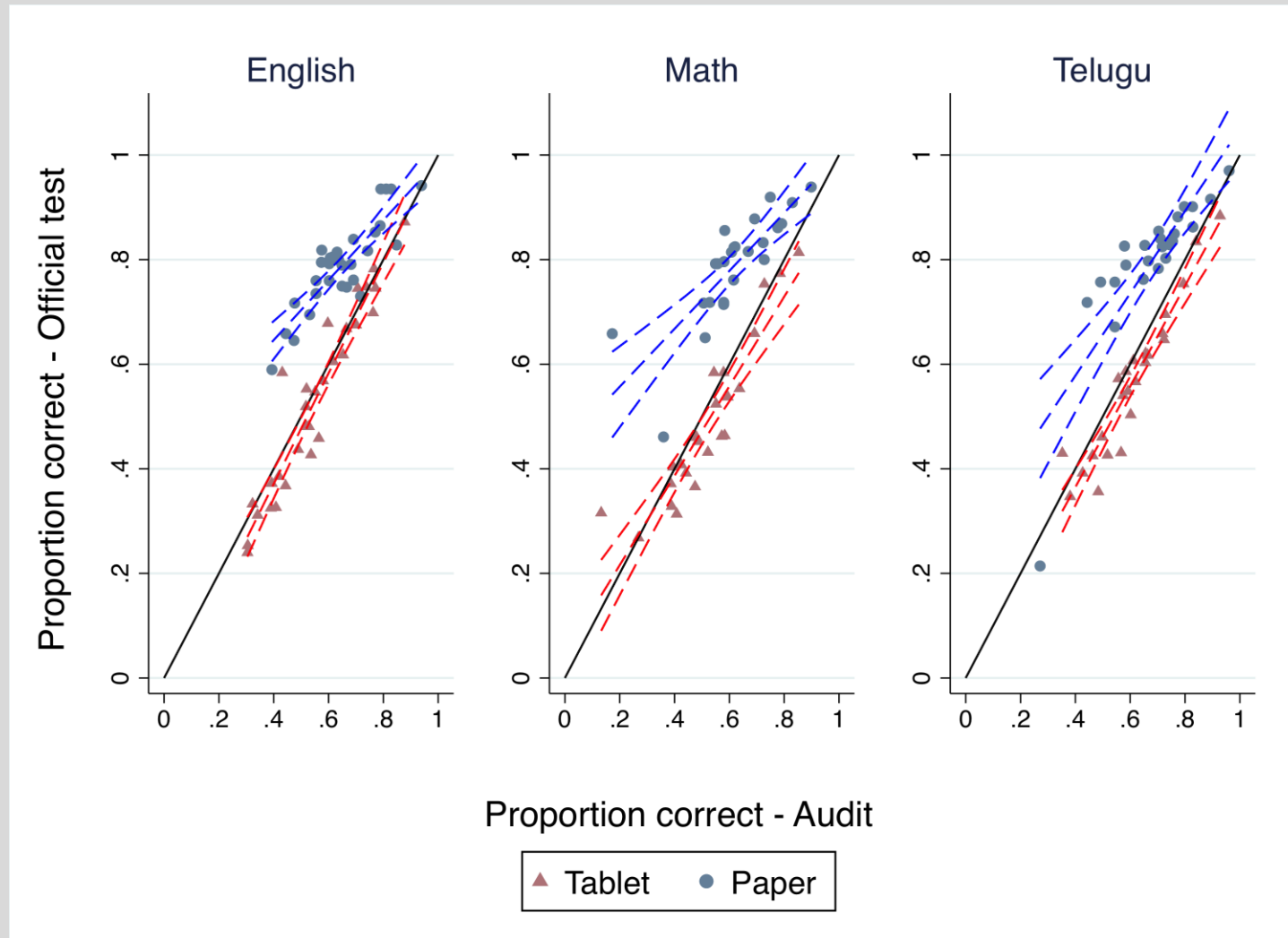
The principal secretary of school education in a central Indian state faces a conundrum—elated at being placed in the top quartile in India in MHRD’s annual National Achievement Survey (NAS), he has been replaced with a sense of despondency at being rock bottom in the recent ASER 2018 results. The secretary has the state’s own learning data collected randomly across 1 lakh students in government schools—which is different from NAS and ASER. There is also a file on learning outcome data from Aspirational Districts collected by the NITI Aayog that is as yet unread on his desk—his fear that it will paint a fourth picture is not unfounded.

As a consequence of conflicting measurements and lack of quality data, the objective of index-based measurement systems—to prioritise and identify weaknesses for improvement—is largely a lost cause. Today, if any bandwidth is spent in states, it is on wondering “what really is the truth.” Data-based learning assessments and rankings that should have been a clarion call to action for states have degraded into a source of frustration and cynicism, as well as the target of ridicule.

(Financial Express, Feb 28, 2019)

# It may be possible to reduce manipulation

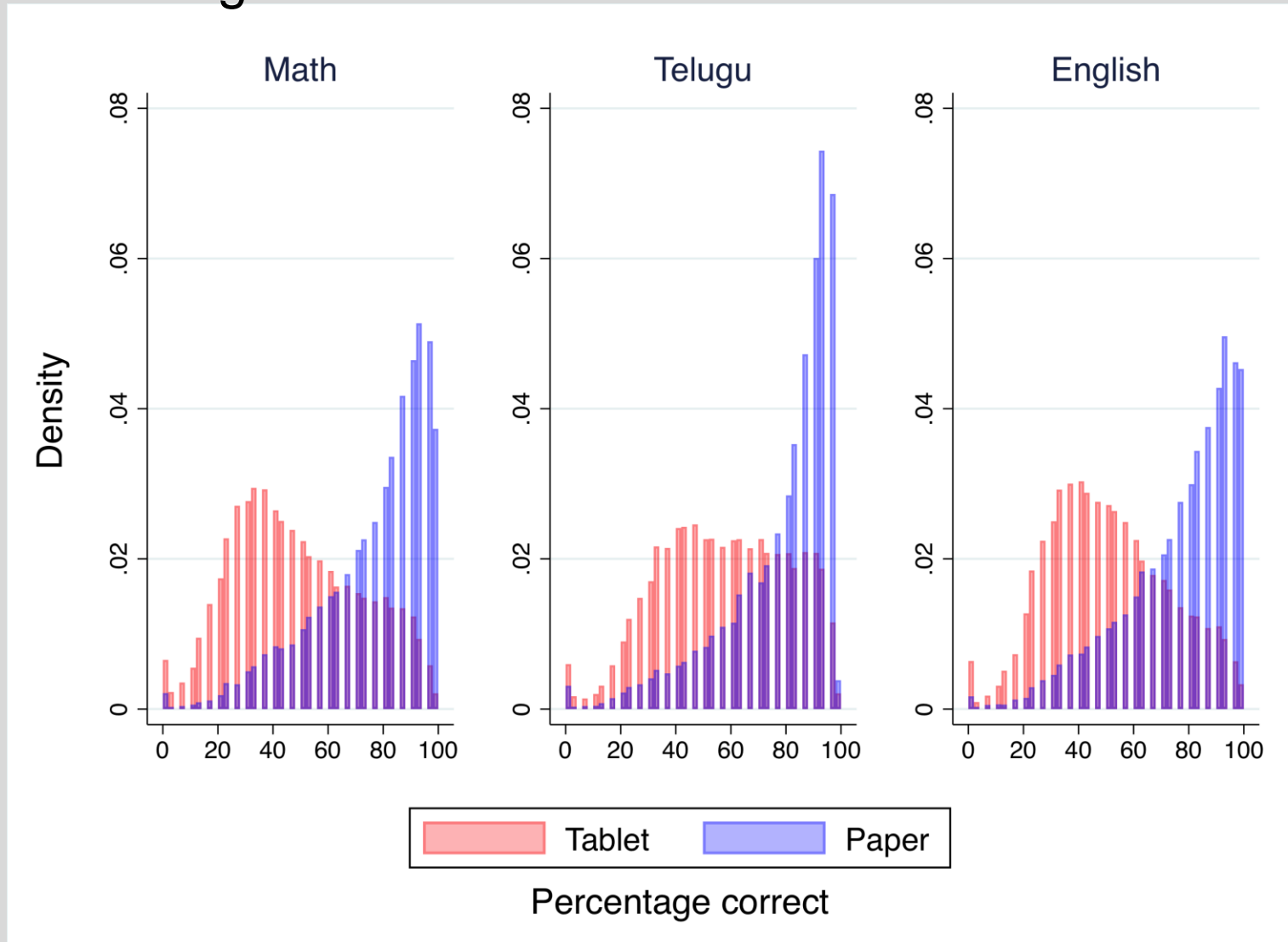
## Tablet-based testing in Andhra Pradesh



Singh, A (2023) Improving administrative data at scale: Experimental evidence on digital testing in Indian schools, Conditionally Accepted, *Economic Journal*

# It may be possible to reduce manipulation

## Tablet-based testing in Andhra Pradesh



Singh, A (2023) Improving administrative data at scale: Experimental evidence on digital testing in Indian schools, Conditionally Accepted, *Economic Journal*

Child-level Panel Data:

The Young Lives Study

# Background

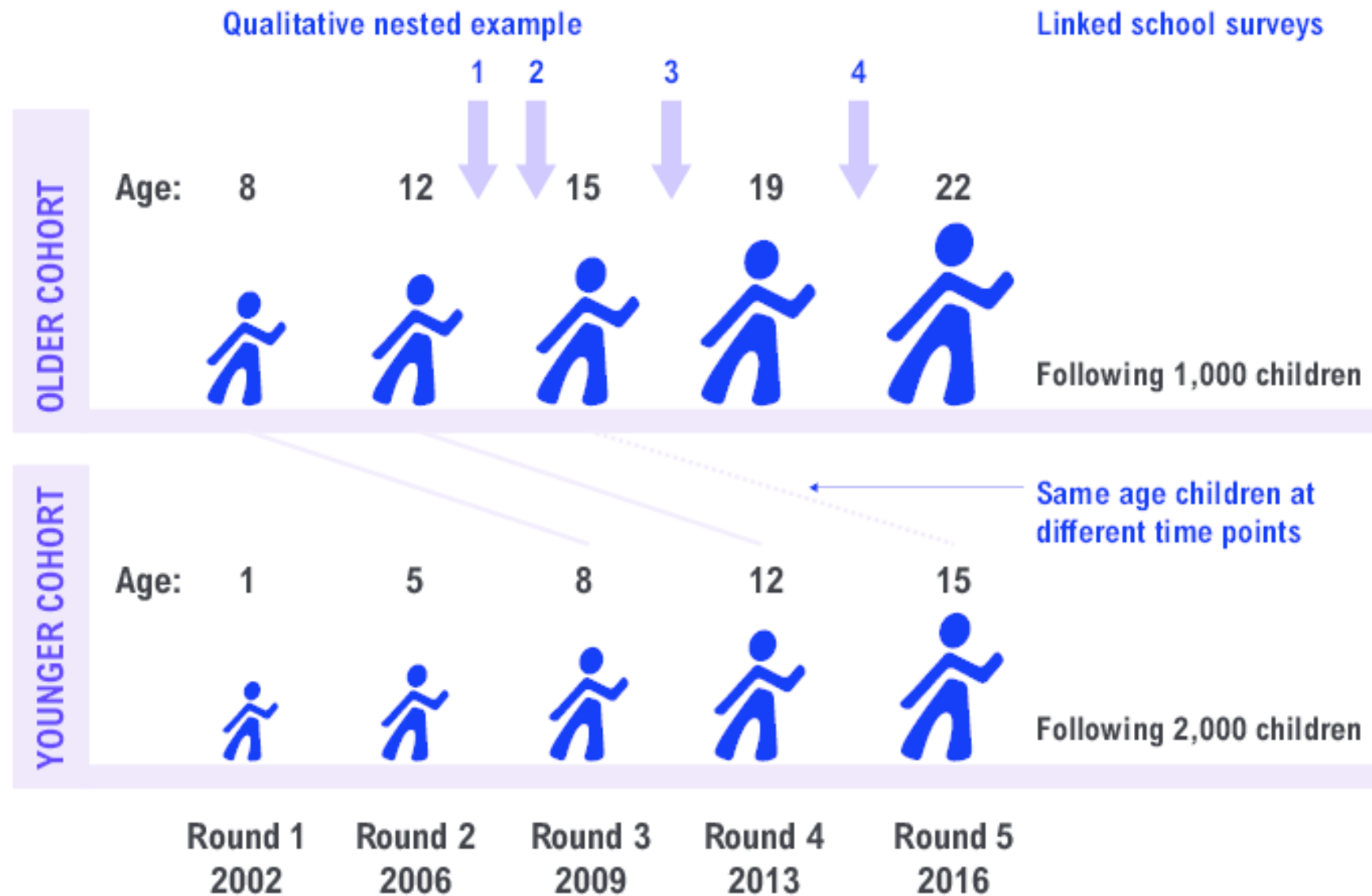
- So far, everything has been about India-wide cross-sections
  - Those are what we need to measure national levels and trends
  - Agenda-setters as to what is important
- But a separate use case is data for research
  - Causal effects of education inputs, long term associations etc.
  - Much of the most influential research on education in India is based on primary data collection
  - That requires resources, also clarity in questions, survey design, randomization (often)
  - I won't be talking about that but rather what can students and researchers in India access easily




# The Young Lives study

- Two cohorts of (once) children born in 1994/95 and 2000/01
  - Surveyed 5 times since then
- Four countries: Ethiopia, Peru, Vietnam and two states in India (AP and Telangana)
  - Not statistically representative of AP and Telangana but with broad range of variation
  - Good for studying panel dynamics, less good for figuring out levels
- Also accompanying qualitative studies, school surveys in some countries/years
- **Best panel data in India to study many educational issues**
  - Data available freely with extensive documentation ([www.younglives.org.uk](http://www.younglives.org.uk))


Five rounds of data collection in four countries:  
Ethiopia, India (Andhra Pradesh and Telangana), Peru, Vietnam



# What have I used it for?



Journal of Development Economics  
Volume 113, March 2015, Pages 16-32



Private school effects in urban and rural India: Panel estimates at primary and secondary school ages

[Abhijeet Singh](#)

Show more ▾



Journal of Public Economics  
Volume 206, February 2022, 104570



Short communication


Test scores and educational opportunities: Panel evidence from five low- and middle-income countries ☆

[Jishnu Das](#)<sup>a</sup>  , [Abhijeet Singh](#)<sup>b</sup>, [Andres Yi Chang](#)<sup>c</sup>

## JOURNAL ARTICLE

### Learning More with Every Year: School Year Productivity and International Learning Divergence



[Abhijeet Singh](#) 

*Journal of the European Economic Association*, Volume 18, Issue 4, August 2020, Pages 1770–1813, <https://doi.org/10.1093/jeea/jvz033>

## THE UNIVERSITY OF CHICAGO PRESS JOURNALS

School Meals as a Safety Net: An Evaluation of the Midday Meal Scheme in India

Author(s): [Abhijeet Singh](#), [Albert Park](#) and [Stefan Dercon](#)

Source: *Economic Development and Cultural Change*, Vol. 62, No. 2 (January 2014), pp. 275–306

# Pros and cons

- Pros:
  - Very rich data, freely accessible, extensive documentation
  - Long-run panel, comparable across countries
  - Ideal for students who want to work on issues for MSc, PhD
- Cons:
  - Two states only, two cohorts only, 20 mandals
  - Whether you can answer a policy q, depends entirely on cohort/time
  - Many “obvious” questions are answered already
- **This is a very good starting point for research students, not for data journalists or policy**

Where are we headed?

Speculative notes

# High quality administrative data: the final frontier?

- States and the national government have spent substantial resources and effort in creating an infrastructure for administrative data
  - Many states are moving towards student-level databases
  - States are moving towards large-scale assessments
  - NEP is creating an independent testing agency (PARAKH)
  - Schools already are in the database(s)
- **We already have the technical infrastructure to have world-class digital infrastructure for education and policy research**
- **Two big questions:**
  - Will it be reliable?
  - Will micro-data be made available to researchers and general public?
- These are solvable problems but need support and political will

Thanks for listening