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)

• 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

<table>
<thead>
<tr>
<th>Sections</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>School Profile (Location, Management, Medium of Instruction etc)</td>
</tr>
<tr>
<td>Section 2</td>
<td>Physical Facilities, Equipment, Computer and Digital initiatives</td>
</tr>
<tr>
<td>Section 3</td>
<td>Teaching and Non-Teaching Staff</td>
</tr>
<tr>
<td>Section 4</td>
<td>Enrolment and Repeaters</td>
</tr>
<tr>
<td>Section 5</td>
<td>Incentives and Facilities Provided (Only for Govt. and Govt. Aided Schools)</td>
</tr>
<tr>
<td>Section 6</td>
<td>Annual Examination Results</td>
</tr>
<tr>
<td>Section 7</td>
<td>Board Examination Result</td>
</tr>
<tr>
<td>Section 8</td>
<td>Receipts and Expenditure</td>
</tr>
<tr>
<td>Section 9</td>
<td>Vocational Education under NSQF at Institutional level</td>
</tr>
<tr>
<td>Section 10</td>
<td>PGI and Other Indicators (Only For Government and Government Aided Schools)</td>
</tr>
<tr>
<td>Section 11</td>
<td>School Safety</td>
</tr>
<tr>
<td>Section 12</td>
<td>Gifted Children</td>
</tr>
</tbody>
</table>
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)
What have people used it for? (Kingdon, 2020)

<table>
<thead>
<tr>
<th>Total number of pupils in the school as a whole:</th>
<th>Number of Schools</th>
<th>Percentage of total govt. schools</th>
<th>Number of Teachers</th>
<th>Total Enrolment</th>
<th>Average pupils per school</th>
<th>Pupil teacher ratio</th>
<th>Teacher Salary Expenditure (Rs. Crore)</th>
<th>Govt. Annual Per-pupil Salary Exp. (Rupees)</th>
<th>Govt. Monthly Per-pupil salary Exp. (Rupees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>4,425</td>
<td>0.43</td>
<td>14,304</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>503</td>
<td>4,88,101</td>
<td>40,575</td>
</tr>
<tr>
<td>5 or Less</td>
<td>8,675</td>
<td>0.84</td>
<td>21,277</td>
<td>15,333</td>
<td>1.8</td>
<td>0.7</td>
<td>748</td>
<td>1,27,530</td>
<td>10,628</td>
</tr>
<tr>
<td>10 or Less</td>
<td>21,008</td>
<td>2.03</td>
<td>42,843</td>
<td>1,18,166</td>
<td>5.6</td>
<td>2.8</td>
<td>1,507</td>
<td>52,750</td>
<td>4,397</td>
</tr>
<tr>
<td>20 or Less</td>
<td>71,189</td>
<td>6.87</td>
<td>1,38,033</td>
<td>9,20,254</td>
<td>12.9</td>
<td>6.7</td>
<td>4,855</td>
<td>13,670</td>
<td>7,774</td>
</tr>
<tr>
<td>50 or Less</td>
<td>3,13,169</td>
<td>30.24</td>
<td>6,33,323</td>
<td>95,10,902</td>
<td>30.4</td>
<td>15.0</td>
<td>22,277</td>
<td>47,654</td>
<td>3,972</td>
</tr>
<tr>
<td>All govt. schools</td>
<td>10,35,602</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>6,714</td>
<td>0.64</td>
<td>11,791</td>
<td>0</td>
<td>0.0</td>
<td></td>
<td>677</td>
<td>5,04,256</td>
<td>42,021</td>
</tr>
<tr>
<td>5 or Less</td>
<td>14,991</td>
<td>1.44</td>
<td>26,043</td>
<td>29,638</td>
<td>2.0</td>
<td>1.1</td>
<td>1,495</td>
<td>1,92,539</td>
<td>16,045</td>
</tr>
<tr>
<td>10 or Less</td>
<td>36,365</td>
<td>3.48</td>
<td>68,586</td>
<td>2,04,421</td>
<td>5.6</td>
<td>3.0</td>
<td>3,936</td>
<td>93,285</td>
<td>7,774</td>
</tr>
<tr>
<td>20 or Less</td>
<td>1,16,307</td>
<td>11.14</td>
<td>2,38,213</td>
<td>14,65,423</td>
<td>12.6</td>
<td>6.2</td>
<td>13,670</td>
<td>47,654</td>
<td>3,972</td>
</tr>
<tr>
<td>50 or Less</td>
<td>4,17,193</td>
<td>40.00</td>
<td>9,85,051</td>
<td>1,18,59,775</td>
<td>28.4</td>
<td>12.0</td>
<td>56,529</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All govt. schools</td>
<td>10,43,939</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: www.statereportcards/rawdata/201011 Data analysed here are for 21 major states (counting Telengana as a separate state).
Sanitation and Education†

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)

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

Anjali Adukia† Sam Asher‡ Kritarth Jha§ Paul Novosad¶ Brandon Tan∥

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

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 a, Rajshri Jayaraman b, *

a IIM Calcutta, India
b 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
State-led assessments are manipulated
Pratibha Parv in Madhya Pradesh

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

Qualitative nested example

Linked school surveys

OLDER COHORT
Age: 8 12 15 19 22
Following 1,000 children

ROUND 1
Round 2
2002 2006

ROUND 3
Round 4
Round 5
2009 2013 2016

Following 2,000 children

SAME AGE CHILDREN AT DIFFERENT TIME POINTS

YOUNGER COHORT
Age: 1 5 8 12 15

Same age children at different time points

Following 1,000 children
What have I used it for?

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

Abhijeet Singh

Show more

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

Ishma Das, Abhijeet Singh, Andres Yi-Chang

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

Abhijeet Singh

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

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

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