Towards Enhancing
Data Utility and Accessibility for Schools

Acknowledgments

Pacta is a social sector exclusive law firm and policy think tank bridging gaps in law and policy implementation through rigorous research.

Suggested Citation: "Towards Enhancing Data Utility and Accessibility for Schools, Pacta, Bengaluru, June 2024."

Glossary of Terms

ASER	Annual Status of Education Report
BPL	Below Poverty Line
BRC	Block Resource Centre
CRP	Community Resource Person
CSO	Civil Society Organization
CSR	Corporate Social Responsibility
CWSN	Children with Special Needs
DPEP	District Primary Education Programme
EMIS	Education Management Information System
GER	Gross Enrollment Ratio
HM	Head of the school
KSEEB	Karnataka School Examination and Assessment Board
PICME	Pregnancy and Infant Cohort Monitoring and Evaluation (implemented in Tamil Nadu)

NAS	National Achievement Survey
NRHM	National Rural Health Mission
NIC	National Informatics Centre
NIEPA	National Institute for Education Planning and Administration
NPSSE	National Programme on School Standards and Evaluation
OASIS	Online Affiliated Schools Information System (used by CBSE schools)
PTR	Pupil Teacher Ratio
RTE	Right to Education
SATS	Student Achievement Tracking System
SECI	Socialisation, Externalization, Combination, and Internalization (a model of education developed in Japan)
SSEF	School Standards Evaluation Framework
UDISE	Unified District Information System for Education

Table of Contents

Part I. Introduction

A. Executive Summary

This report is based on exploratory research undertaken under the aegis of Aapti Institute's call for research proposals on the theme of Strengthening Data Ecosystems in Indian Schools. We conducted this research study to map out data submitted by schools to government mandated databases such as (but not limited to) the Unified District Information System for Education Plus ["UDISE/UDISE+"], towards identifying pathways that enables the usage of such data to serve students, shaping better schooling outcomes, moving away from data as a means of oversight and punishment and towards data as a tool for empowerment.

Through secondary research and expert interviews, we found nuanced insights on challenges such as lack of the utility of data for schools, a lack of trust among the stakeholders in the ecosystem and several technical barriers in collecting and reporting this data. We identified power hierarchies in the way data is collected and used, resulting in the observation that reliance on data-informed decisions remains restricted (at best) to policy making corridors or funding decisions for civil society initiatives, and has not permeated to use at schools. We found evidence of inequitable data utility and accessibility for participants such as schools and school teachers (data workers) low in the data supply chain, than for participants such as public administrative instrumentalities that are higher placed in the data supply chain.

We concluded that efforts are needed to enhance the accessibility and utility of data that is collected and reported from the school level - reduce the burden on the school staff, improve internet software and hardware infrastructure and refrain from blaming the school (teacher) for insights that reflected poor learning outcomes/ experiences of students. We concluded with a skeletal framework of indicators that can be used to measure accessibility and utility of the above government databases. We call this the Data Accessibility and Utility Framework ["DAU"].

In the modern context, participative (also known as participatory) design of data collection and reporting data-bases has been shown as being is crucial for mass uptake and adoption by all stakeholders.[1] However, the role of teachers in participatory design of education databases and tools has been limited. Hence, the feasibility of such a system remains to be determined.

---

[1] Tuhkala A, "A Systematic Literature Review of Participatory Design Studies Involving Teachers" (2021) 56 European Journal of Education 641 <https://doi.org/10.1111/ejed.12471>

Six schools and 10 field practitioners were interviewed through semi-structured interviews. Further details of the practitioners and schools who participated in the interviews is available under Table 1 and Table 2 below

Table 1: Practitioner Participant Details

Practitioner	Location	Types of the organization/work
P1	Karnataka	CSO helping marginalised children with their education journey
P2	Tamil Nadu	CSO developing and implementing ed-tech solutions with particular focus on foundational learning
P3	Haryana	CSO social entrepreneur working in grass-roots level education
P4	Maharashtra	CSO to create inclusive classroom experiences for students
P5	Jammu and Kashmir	CSO designing and imparting context-based curriculum in conflict areas
P6	Haryana	CSO implementing classroom observation tools in local schools
P7	Delhi	Educational Policy Expert
P8	Delhi	CSO developing capacity building programs for government school teachers
P9	Assam	NPO engaged in grassroots level work in education and health
P10	Karnataka	NPO that engages with BRCs, CRPs and government schools

Table 2: School Participant Details

School	Location	Type
S1	Uttarakhand	Private unaided
S2	Tamil Nadu	Affordable Private
S3	Karnataka	Affordable Private (NGO Run)
S4	Uttar Pradesh	Affordable Private
S5	Kerala	Private Unaided
S6	Assam	Affordable Private (NGO Run)

B. Methodology

B1. Research Overview:

The study was conducted using qualitative methods. As part of the process, an exhaustive desk review was undertaken, followed by in-depth semi-structured interviews with 16 stakeholders. Interviews were conducted from 7 September, 2023 to 25 October, 2023. Thematic analysis was used to unveil recurring patterns within the data.

B2. Interview process

Stakeholders were categorised into two groups: policy/ field practitioners (in school education) and schools. Interview respondents were contacted through Pacta's and Sphoorti's networks. Six schools and 10 field practitioners were interviewed through semi-structured interviews. Out of these, we visited three schools in person, and two ran us through the databases they reported to. The other schools and all the practitioners were interviewed through video conference calls. The interviews were recorded after taking consent and were later transcribed using Otter.ai. The schools were located in Karnataka, Tamil Nadu, Uttar Pradesh, Uttarakhand, Kerala and Assam. The field practitioners included people working with Civil Society Organizations ["CSOs"] at policy research or implementation levels. We undertook thematic analysis to discern recurrence of patterns in the collected data. The transcriptions were coded, and the codes were systematically categorised to derive themes. Sub-themes were identified within each overarching theme.

B3. Analysis of data and development of DAU

The **'World Bank's Framework for Assessing the Quality of Education Statistics'** served as a guiding framework for developing DAU parameters. Some of the indicators listed in this framework were tailored to suit the needs of the Indian educational datasets, while some indicators have been included upon analysis of the findings of this Report.

B4. Objectives:

This Report focuses on three key objectives:

1) To fill in the gaps in literature on the accessibility and utility of data to schools

2) To understand the current attitudes towards data in the school ecosystem

3) To encourage and enable schools to take more data-driven decisions

B5. Gap in Literature

There is a noticeable gap in the existing body of literature on the utility of educational databases in India. To ascertain the extent of this gap, we used VOSViewer, a mapping software. Our analysis showed that there is a dearth of literature in the education sector in the context of data use. A detailed description of the method used to derive these maps has been attached as Annexure I.

We also found that there was limited association in existing literature between "UDISE" and "decision making", "outcome", "analytic" etc in India.

In Figure 1 where although the points representing "data" and "education" are situated relatively close, the point representing "education sector" appears at quite a distance from "data". In Vos viewer, the application used to generate Figures 1-3, the closer the keywords appear, the more the two are associated with each other in the literature.
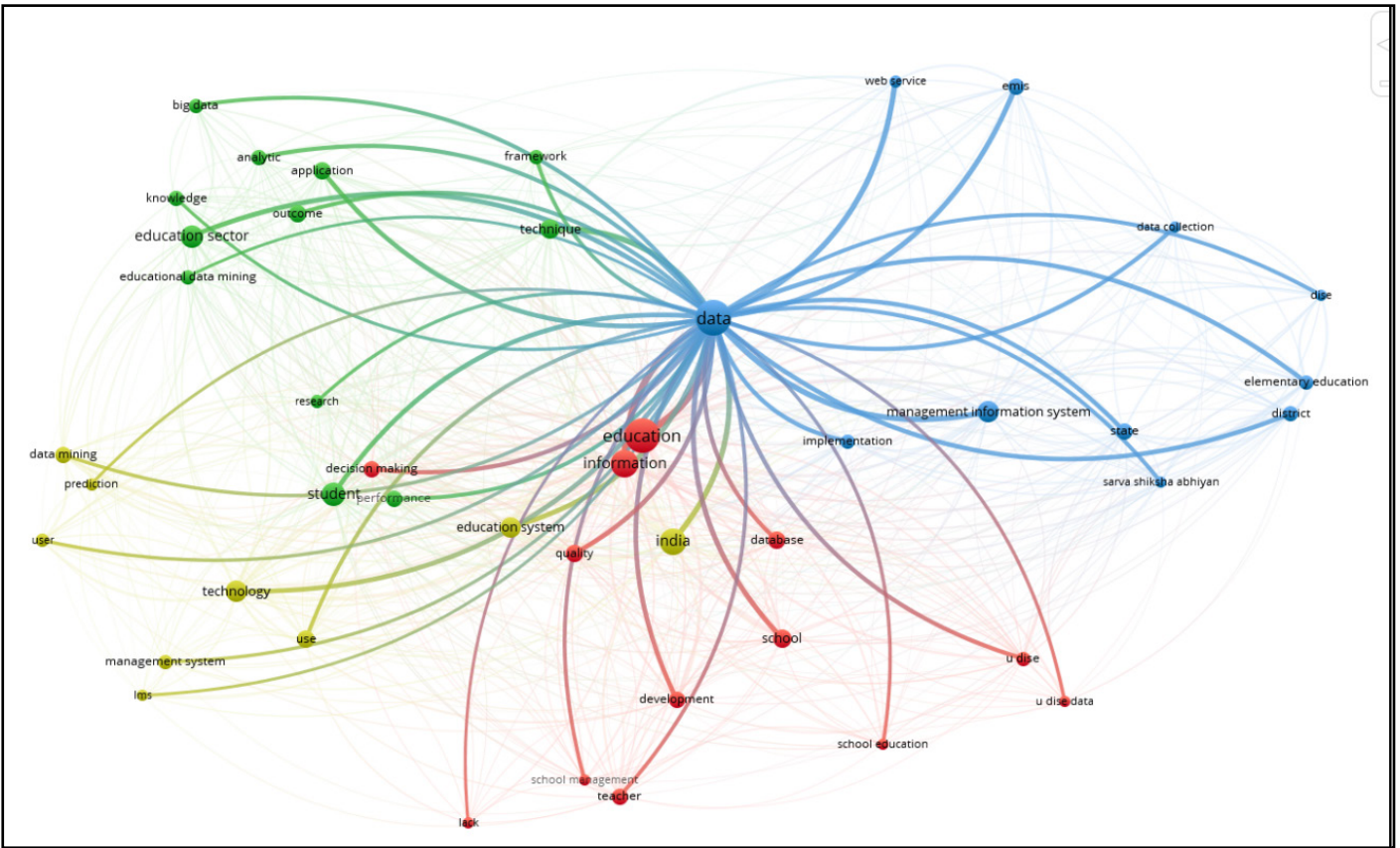
Figure 1: Relation of "data" with other keywords in the existing literature
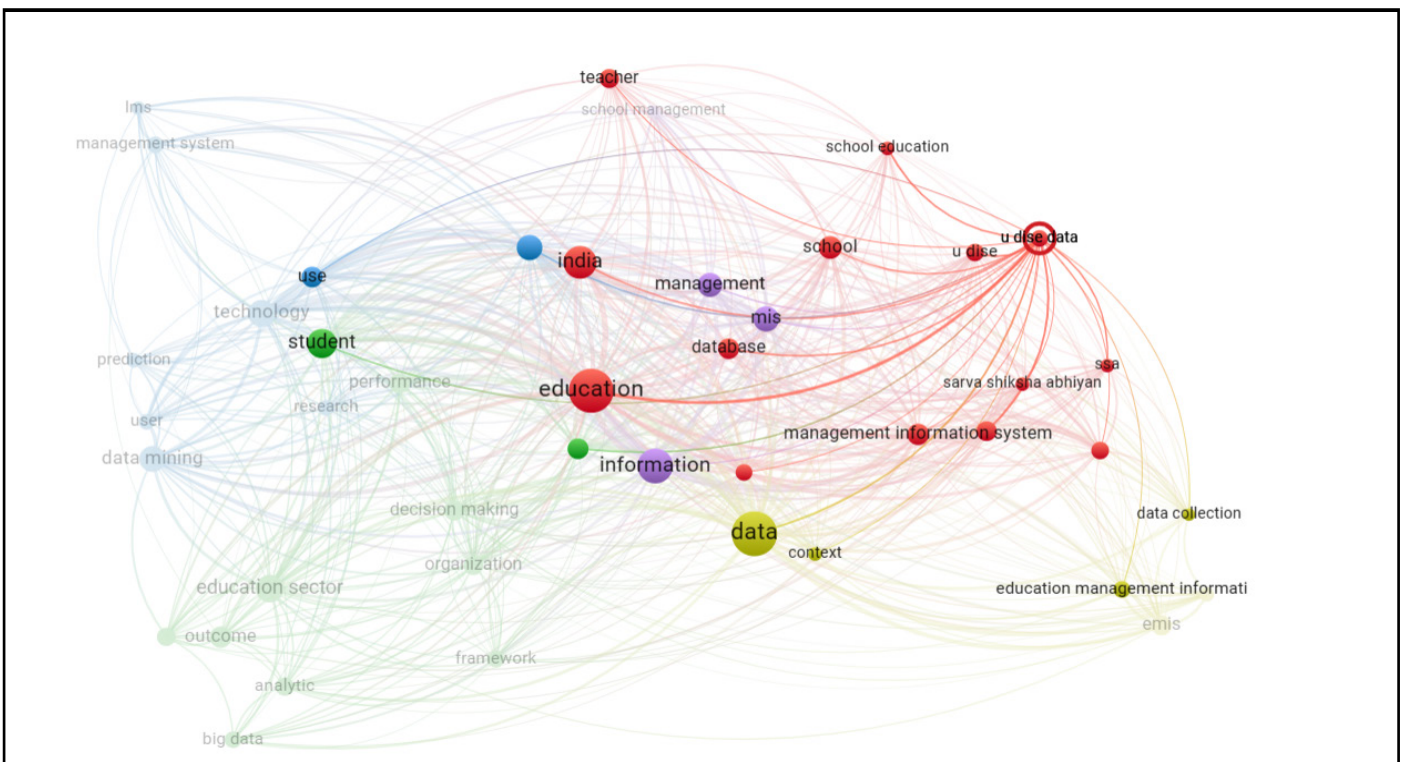
Figure 2: Relation of "UDISE data" with other keywords in the existing literature
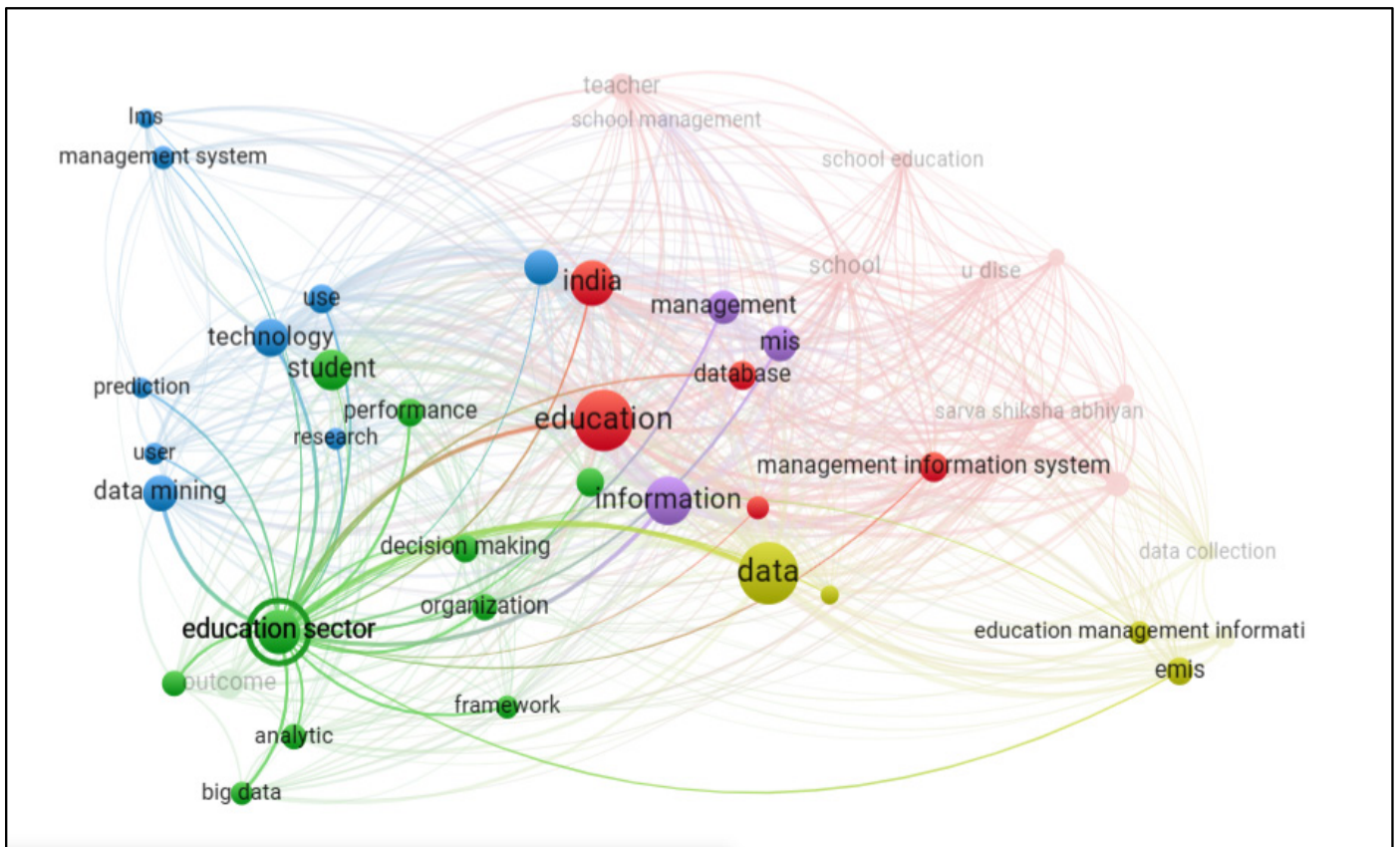
Figure 3: Relation of "education sector" with other keywords in the existing literature

Insights from the Literature Review conducted using VoS Viewer

The mapping exercise revealed a disjointed relationship between the education sector and data as they belong to different clusters in the map. This implies that the role of data in the education sector is not well established in the existing literature. Similarly, the nodes, **"education sector"** and **"EMIS" (Education Management Information System)** are distant from each other. However, the relationship of **"education sector," "decision making," "big data"** shows the emerging relevance of such concepts in the education sector.

⚠️ B6. Limitations of the Study

1. Schools had apprehensions about data-related conversations, hence we had a limited sample size

The Research team faced difficulties in stakeholder outreach: both public and private schools hesitated to participate in our interviews as they were uncomfortable speaking to us about data. Often despite our reassurances, they assumed we were looking for school data. We were able to successfully interview only six schools.

We tried to balance this out by speaking to schools from different states and also getting many interviews with experts to get a more diverse set of perspectives.

2. Schools do not see the utility of UDISE and State EMIS data

Due to the absence of data culture in schools, schools had not previously thought about the utility of government data to generate insights. Many of them expressed the same during their interviews. Schools were also uncomfortable speaking to us about consent, security and data protection measures. Since schools are not equipped with the skill-sets to put data to use, this study might be ahead of its time and focus could instead be on creating a data culture before finding out what data insights schools find useful.

3. Data collection practices are fragmented

Different schools had adopted different measures to collect and maintain the data required for mandatory reporting. Some states have state-level databases. Although UDISE+ is a mandatory for recognized schools, the S2, a school we interviewed in Tamil Nadu, was not aware of UDISE+. Some research participants preferred to remain anonymous, so we have anonymized all the quotes from our qualitative interviews to maintain consistency.

Part II. Mapping Educational Data Reporting & Utility for Schools

This Part delves into the data collected by Indian schools for internal purposes and as part of their reporting requirements for mandates by the Central Government, State Government, funders etc. This Part also discusses how different stakeholders use the data.

A. Central Level Reporting Mandates

UDISE

The school education system in India is one of the largest in the world. Across 28 states and eight Union Territories, there are over 1.48 million schools, over 9.5 million teachers and over 265 million students.[2] These schools all comprise students from a multitude of socio-economic backgrounds, and require robust and timely information collection. This information is essential in the creation of effective policy interventions to improve education outcomes for children across the country.[3]

---

[2] 'UDISE_Booklet.Pdf' <https://dsel.education.gov.in/sites/default/files/update/UDISE_Booklet.pdf > accessed 23 September 2023.

[3] UDISE+' <https://udiseplus.gov.in/#/page/about> accessed 20 September 2023.
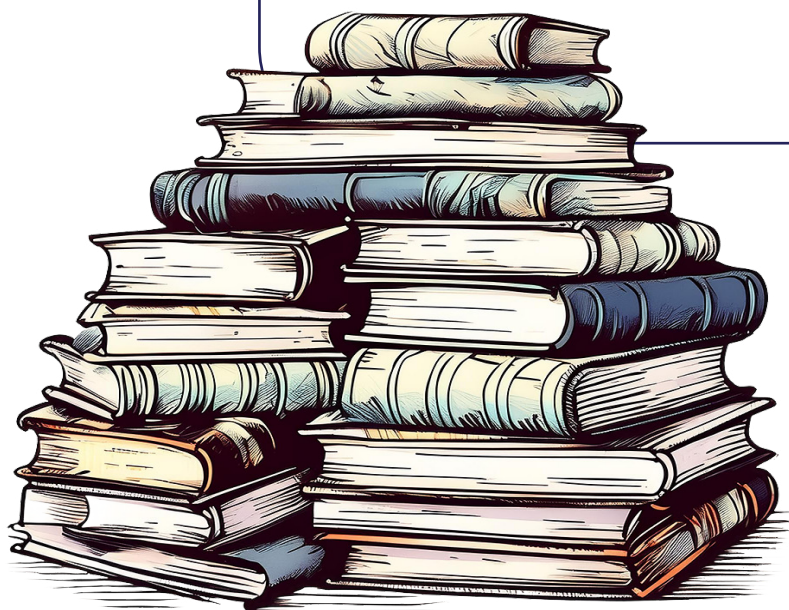
One of the ways in which educational information can be collected to improve these outcomes, is by using EMIS **["Educational Management Information Systems"].** These systems are used to systematically gather, understand and analyse data to support future education outcomes.[4] In the early 1990s, one such system was developed under the DPEP **["District Primary Education Programme"].** This computerised information system was created for Classes I to VIII and was called DISE **["District Information System for Education"].** In 2008-09, a separate system was created for Classes IX to XII. This was called SEMIS **["Secondary Education Management Information System)"].** Due to the fissured nature of these systems, it necessitated a system that integrated them. This system, which was introduced in 2012-13, was called UDISE **["Unified District Information System for Education"].**[5]

The primary objective of the UDISE was the comprehensive collection of information pertaining to schools' infrastructure, enrolment, pupil-teacher ratios, and other such parameters. This was done to measure progress and identify gaps within the education system in India. This data was then used to make involved stakeholders make more sound decisions. Creating a centralised management information system was done with the aim of making data accessibility and improvement easier and more efficient. However, it came with a host of limitations and challenges; the long time taken due to offline collection and entry made it outdated, the Ministry had to deal with two sets of data from some states since some of them preferred using their own EMIS, the data entered was not verified, and the contractual and outsourced nature of the MIS officers entailed that there was no accountability for unreliable data.[6]

With the aim of correcting the limitations found in the UDISE, the UDISE+ was then introduced in 2012-13. It aimed to offer a user-friendly interface and more immediate utility for collected data.[7] This data was also hosted within the secure space of the NIC **["National Informatics Centre"],** thereby enabling efficient storage of data. Now functioning as the most comprehensive governmental source of information for the school education system in India, it serves as data of immense potential and importance.

While individual states can have their independent EMIS, the mandate of the UDISE+ facilitates a system where there is an aggregate of information on the central level. This can thus be used in the formulation of budget plans, policy decisions, and other measures. The UDISE+ accounts for nine different individual metrics, with multiple sub-categories. All schools that we interviewed reported to UDISE+, though with diverse experiences in reporting and utilising the data, as outlined in Parts B and C of this Report. **In the words of P2, a field practitioner from Tamil Nadu,**

> "
> Some states are far ahead. Some states are are comparatively behind in tech readiness but everyone has UDISE, that's the critical part.
> "

4 Framework_SABER-EMIS.Pdf' <http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/Background/EMIS/Framework_SABER-EMIS.pdf> accessed 20 September 2023.
5 'UDISE_Booklet.Pdf' <https://dsel.education.gov.in/sites/default/files/update/UDISE_Booklet.pdf > accessed 23 September 2023.
6 'UDISE_Booklet.Pdf' <https://dsel.education.gov.in/sites/default/files/update/UDISE_Booklet.pdf > accessed 23 September 2023.
7 'UDISE_Booklet.Pdf' <https://dsel.education.gov.in/sites/default/files/update/UDISE_Booklet.pdf > accessed 23 September 2023.

NAS

National Achievement Survey also known as NAS is a national level assessment of student learning outcomes in India for grades 3,5,8 and 10. NAS is conducted once every three years; the first report was released in 2017 and the second was released in 2021. The survey in 2021 records the adverse effects of the COVID-19 pandemic on learning levels of students in India. P2 observed,

However, researchers who studied the accuracy of NAS data concluded that NAS data could not be relied on as (1) the state averages of NAS were **"artificially high"** and (2) the relative performance of states cannot be measured. The researchers observed that NAS was conducted with a lot of thought and the issue is attributed to the intrinsic difficulties in collecting accurate learning outcome data especially through assessments of schools.[8]

> "
> UDISE also does not collect student outcome data religiously. We depend on NAS or equivalent for student outcome data.
> "

P4, a field practitioner based in Maharashtra felt that,

> "
> the sample statistical methods that are being used at a national level by NAS to get this data points,they are sound and solid, there is no discrepancy in that actually....More or less, the data we collect for our CSO aligns with the NAS data.
> "

---

[8] Johnson D and Parrado A, "Assessing the Assessments: Taking Stock of Learning Outcomes Data in India" (2021) 84 International Journal of Educational Development 102409 https://doi.org/10.1016/j.ijedudev.2021.102409

NPSSE

The NPSSE (National Programme on School Standards and Evaluation), more commonly known as the Shaala Siddhi was introduced in 2015, and addresses the need for improved school performance in the Indian education system. **"School Evaluation as the means and School Improvement as the goal"** is the initiative's motto.[9] Launched by the NIEPA (National Institute of Educational Planning and Administration), the NPSSE is structured on the belief that each individual school and its performance must be continuously evaluated, with its broad objective being to **"develop strong conceptual frameworks in education."** Now, it aims to reach 1.5 million schools in the country and its own framework called SSEF (School Standards and Evaluation Framework) has been developed.[10]

NPSSE differs from other mechanisms of evaluation in the sense that schools are viewed as institutions capable of assessing and reforming themselves. By removing external factors from the equation of improvement, the school in itself is viewed as an organisation capable of reforming itself.[11] NPSSE does not look at self-evaluation exclusively as an exercise of identifying weaknesses or deficiencies. Rather than focusing on compliances, it acknowledges the potential of schools to improve themselves, it encourages schools to critically analyse their own strengths and shortcomings to assess the areas for improvement.[12] Reporting requirements under the NPSSE are classified and organised under seven **'key domains,'** which are significant criteria for the evaluation of schools.

Among our interviewees, only one school had been using the NPSSE Database. Even this school complained about the arduousness of the process, and was unaware that reporting to NPSSE was voluntary.

> ❝
> The UDISE training person had informed us that filling in NPSSE was mandatory, like the UDISE. NPSSE is a horrible process. We schools have to compute the percentage for everything including attendance, how many students scored above 90 percent etc.
> ❞

P9 also validated the difficulty of reporting to NPSSE.

> ❝
> It is very overwhelming, but we have customised it to use it.
> ❞

[9] J InformationBrochure_English.Pdf' <https://shaalasiddhi.niepa.ac.in/pdf-doc/InformationBrochure_English.pdf> accessed 20 September 2023.
[10] 'ProgrammeDocument_English.Pdf' <https://shaalasiddhi.niepa.ac.in/pdf-doc/ProgrammeDocument_English.pdf> accessed 20 September 2023.
[11] 'School Evaluation | School Improvement | ShaalaSiddhi' <https://shaalasiddhi.niepa.ac.in/about.html> accessed 20 September 2023.
[12] 'ProgrammeDocument_English.Pdf' <https://shaalasiddhi.niepa.ac.in/pdf-doc/ProgrammeDocument_English.pdf> accessed 20 September 2023.

Practitioners brought up the difficulty of filling the database, which could be a reason for its low level of adoption by schools. Two field practitioners mentioned that schools do not report to NPSSE unless there is pressure. As per a practitioner from Delhi, the large number of data points collected and evaluated by NPSSE made it difficult for a single teacher. **He said:**

> " "
>
> It is a pain in your backside [you will lose focus of why you are collecting this even though it is comprehensive.
>
> " "

B. Other Reporting Requirements for Schools

State level databases

Besides UDISE+ mandates, some individual states also have their own reporting mandates. From our interviews, we found that schools report to a state level databases in Delhi, Maharashtra (SARAL), Tamil Nadu (MIS), Punjab, Karnataka (SATS), Haryana (NIPUN), and Assam (Shiksha Setu Ahom). This is not an exhaustive list, and only points towards the databases mentioned by the schools and domain experts. All these state level databases have different levels of efficiency, collect different sets of data, and have varying levels of interoperability with UDISE+. Shiksha Setu Ahom, for instance, is not interoperable with UDISE+, while SATS is. Though they have varying levels of interoperability with UDISE, P2 mentioned that they all **"speak to UDISE."**

> " "
>
> They may not be directly linked but they will speak. They do some extraction etc, and they make [state[data UDISE compliant and they will feed it back into the back end. TC provisions are there, you can track students who transfer.
>
> " "

The approach taken by different states on their state EMIS is different, and the utility that the state governments derive out of it may also vary. **In the words of P2:**

> " Different governments have also approached it differently. Telangana has taken up partnership with NIC. Tamil Nadu has gone with building its own tech team of sorts. Uttar Pradesh has gone with an open tender where people build partnerships, things like that. The fundamental is the state admin's capacity to manage it like any other program that they manage. Ultimately if they are not able to manage it, you can derive very little use of it...because of the concurrent nature of education etc itself, ideal would be to have a national database where each state is able to replicate use and drive their reform agenda, but that seldom is the case. "

Other external databases

S3, a Karnataka school, had additional reporting obligations to the Karnataka School Examination and Assessment Board. S4, a school from Uttar Pradesh also mentioned that they used OASIS, the CBSE school information system, but another CBSE school we interviewed did not refer to this.

Data collected for internal purposes

Five out of six schools that were interviewed also collected and analysed data for various non-state mandated purposes. NGO run schools especially, leverage data effectively to report to their funders. **S3, an NGO run school based in Karnataka found data to be useful for several purposes:**

> " To funders on the type of impact we create, we do some data collection on our end too for internal purposes. We report to our funders on academics and performance of students in extracurricular activities. We also do a family survey, skilling programmes for the parents and health data of children as this is linked to learning outcomes. Analysis is done to assess drop out trends. "

Besides this, schools also collect and maintain data on relatively basic softwares such as Microsoft Excel for administrative purposes, learning outcome analysis, and assessing the effectiveness of teaching. One school also mentioned the use of KREO, a data management software to track attendance.

Research Team Insights on how schools perceive "data"

Thus schools and field practitioners alluded to the burdensome and cumbersome nature of reporting obligations in the context of data collection for state mandated databases. However there was a broad-based acknowledgement that data insights through independent data collection processes were a necessity - both in the context of self-administration of the schools as well in case of reporting needs of funders. Through interviews across schools, the Research Team encountered a guarded approach of interviewees when the word **"data"** was brought up, often necessitating the clarification that the research team was not seeking data from the school, but intended to have conversations around the data collection and reporting experience of schools. This led us to understanding that the word **"data"** in the school education system was associated with punishment and adverse consequences.

C. Utility of Data for Schools

The Research Team has triangulated the insights gathered from interviews, secondary literature and data fields collected by UDISE+ and SATS to identify existing uses of data reported for government mandates. The list of the existing uses compiled by the Team (Refer to Table 3 below) has been classified based on the utility it serves for different stakeholders in the school data ecosystem.

Table 3: Existing uses of data collected by schools as part of reporting mandates

Schools
1. There is some level of analysis at the school level. One practitioner and one school mentioned that they use it to check enrolment rates.   2. Schools use insights from infrastructure related data to inform their fundraising strategy from the community   3. Schools use databases to effectively keep track of students transferred from a different school to theirs or vice versa   Public private partnership schools, run by NGOs use data at a larger scale. They are just not dependent on the costs that the government is allocating. They use data effectively to secure external funding. We spoke to three such schools

Goverments

1. The Government uses it for checking the Right to Education Act (RTE) compliance of the schools. If schools are not found to be compliant, they are shut down.
2. Two practitioners mentioned budgeting as an important use of the UDISE data. In the words of a field practitioner from Chennai, *"The most fundamental and primary use case for the data collected by UDISE is the annual work plan and budgeting that is done by the Ministry of Education and hence their state-wise nodal body called SSA (Samagra Shiksha). The primary use of UDISE data is for funds disbursement."* Another practitioner from Haryana also elaborated on this: *"They broadly use this for the purpose of budgeting. Kitna budget iss baar ka plan karna he, aur kitna budget release karna he? Kitna gaya tha, kitna aaya? ("How much budget has to be planned this time, and how much budget has to be released? How much had gone, how much had come?") - Is it in line with the number of dropout students that we identify at the beginning of the year? So those sorts of utilization, the use of data is limited to that only."*

CSOs Influencing Policy

1. Civil Society organizations use the data for understanding the issues in school education at scale. P2, who works with a CSO, said, *"It helps us understand and quantify some of the challenges that the public education system faces. It also throws light on certain nuances to the issue. For example, people always talk about teacher vacancies, you must have heard that in government schools, there is a lack of teachers etc. But then having the UDISE data in hand, you can very easily calculate the pupil teacher ratio. And you will see that after the RTE Act, many states are compliant at a pupil-teacher-ratio norm level. Which basically means that there is not a shortage of teachers per se at a state level, but distribution is not equal. So what is more critical is probably rationalisation. So I think that's why over the last 5-6 years, at least the informed CSOs talk about rationalisation rather than recruitment."*

Part III. Findings On Data Collection and Utilisation Experiences of Schools

This Part delves into the findings from qualitative interviews with schools and practitioners working closely with them. This Part includes insights on the processes of schools in data collection, the challenges they face in collecting data for reporting requirements and deriving utility from this data. The Part concludes with suggestions on the potential uses that schools can derive from the data collected as part of reporting mandates.

A. Processes and Practices in Data Collection

All six schools interviewed use computers to maintain UDISE+ data. Of the six schools interviewed, three of them also maintain manual records. The divergence in reporting mandates from one state to another led to different views by the schools we interviewed on the time taken to complete data entry.
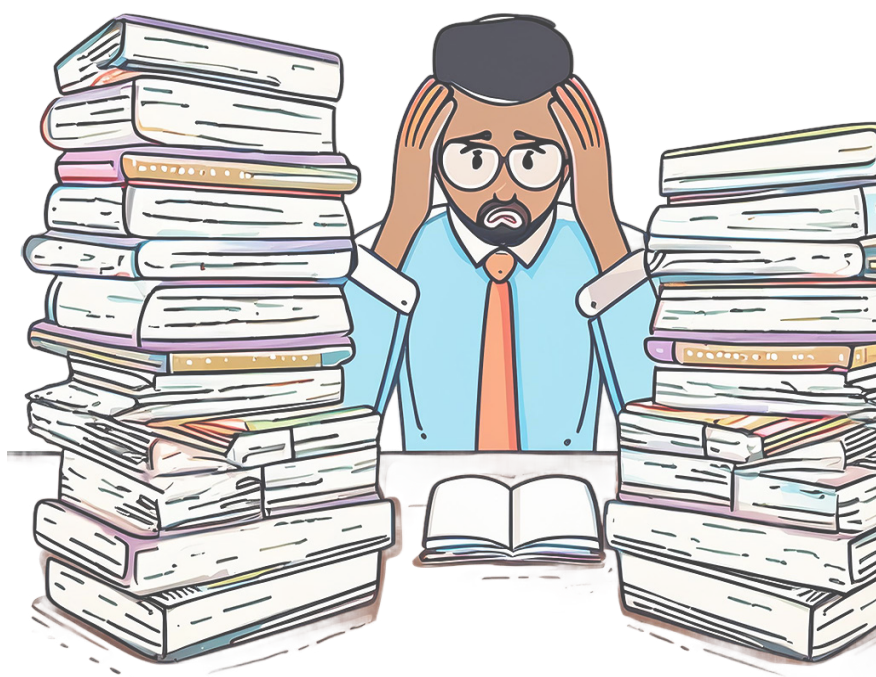
1. Time required for data entry vary

UDISE+

S1, a school based in Uttarakhand reported that it requires 2-3 days for completing the yearly UDISE+. S6, a school in Assam also said that UDISE+ data takes only 10 hours every year. However, S6 reported that Aadhaar related data which is required from 2023-24 for the UDISE + database required about 60 hours of data entry involving multiple teachers.

SATS

S3 said that they required 30 minutes to enter data into SATS for each student who has joined school for Grade 1. Aadhaar verification on SATS takes only a minute per student. However, entering data into the KSEEB database is time consuming- **The person responsible for uploading the data said,**

> "
> I have to upload photographs and signatures with exact specifications of 60x80 kb and 20x50kb. This takes up a lot of time, I spend one week on this.
> "

2. Human resources required for data entry vary

Four schools had one designated person for data entry. The other two schools said they had designated two or more staff who are (usually) computer teachers. Two schools reported that when circumstances arise, other teachers of the school are also asked to help in completing data entry.

3. Consent For Data Collection & Data Security - Awareness and Implementation

Two schools reported that they restricted data entry and access to the government database to only a few teachers for security reasons.

S5 (Kerala) circulated Google forms to parents asking them to fill in the data fields. Five schools reported that parents are informed about the government mandates of data collection but none of them **"specifically take consent."** We requested one of the schools to share the form they circulate to parents which has been produced under Annexure II of this report.

Two schools reported that they get consent from the parents. **S4 stated,**

> "
> They know we tell them in rural set up people do not hesitate to share they are used to giving Aadhar, BPL etc
> "

S5 revealed:

> A message is sent to parents informing them of the data to be collected. Usually, no objections, but some parents have issues sharing Aadhar details; they share it only when informed that they have no choice.

The government databases also raise concerns of data retention. S3 showed us that they could access SATS data of even students who passed out in 2016-17.

Research Team Insights on Consent culture in Indian schools

The research team found that when asked the question of whether schools take consent of parents for sharing the data that gets fed into mandatory databases, schools were aware about the lack of an express consent mechanism. Some schools perceived informing parents on what the data was being requisitioned for as **"consent"**. This also led us to believe that while the UDISE+ data reporting mandate is made on schools, schools may not be equipped with a standardized instrument to record consent for collecting the data.

Further, due to the small size of the sample, details as to exact amount of time spent by school staff in collecting and uploading the data to fulfill mandatory reporting obligations could not be ascertained with certainty. Ethnographic studies would help to bring more nuanced observations surrounding data practices, resources and experiences.

B. Challenges In Data Collection

1. Technical Challenges

1.1. Access to internet is poor

Only about 20% of schools have access to the internet.[13]

Only 30% of schools have a computer.

This means that over 1.8 million schools do not have access to the internet that enables the timely updation of data.
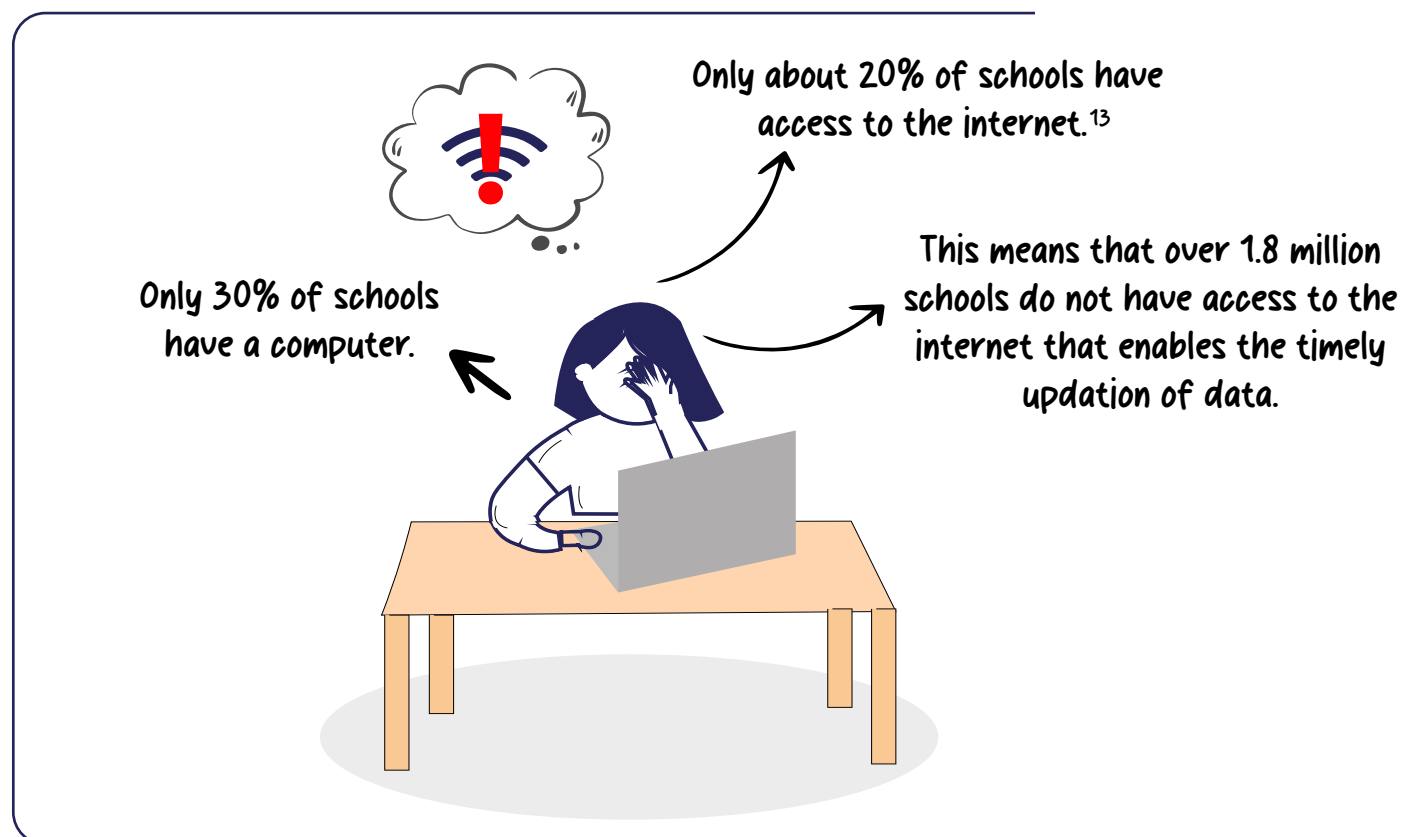
Amongst our interviewees, two schools had manual records of data and due to poor internet, they uploaded this data online when internet was available. This issue is amplified in rural areas and conflict ridden areas of India. P5, **a field practitioner in Kashmir talked about this:**

> Another thing is your relationship with technology first, but also your relationship with the internet. When it's new of course everybody wanted to be on it for the magic of it, then everybody wanted to be on it for the utility of it, and now everybody wants to be on it out of habit. Our children and our teachers are still at the magic stage. Because it's still like- will I have Internet today? So your relationship with the internet is not built enough.

---

[13] UNESCO, Technology in education: A tool on whose terms? (2023 GEM Report) 288.

1.2. Lack of systematic ICT training to teachers

Two Schools reported that training for data collection and entry was given only in the year when the data reporting was made online.
P1 based in Karnataka observed that many schools practised "passing the ownership." In the words of the practitioner:

> passing the ownership– for example HM takes the training but he[she passes the work to another teacher

When asked about how the schools manage in such cases, they replied,

"If the school needs help in uploading data, they have to pay some cyber people out of their own pocket"

1.3. Other Technical Challenges

A field practitioner based in Karnataka succinctly portrayed the multitude of challenges to data collection:

"In some cases, the teachers are trained very well, but if we go to the school there is no computer, if there is computer there is no internet, if there is internet then some system error or app is not working"

Error

Not Responding
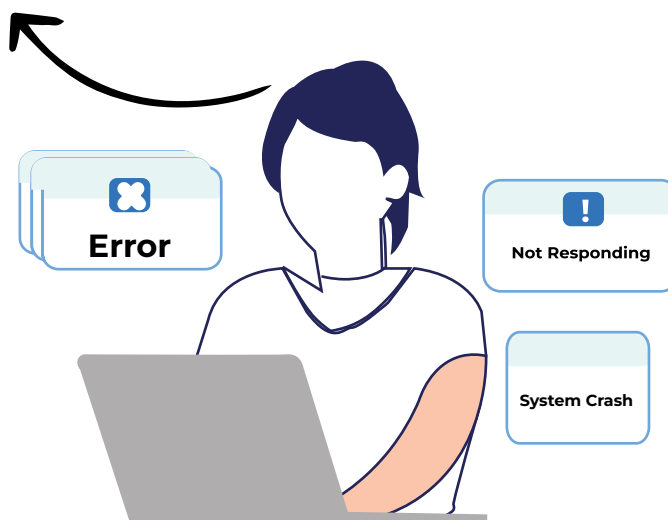
System Crash

S3 and P10 based in Karnataka reported that SATS servers crash regularly. **The field practitioner said :**

The system capacity SATS[is not enough, teachers enter data at 11-12 at night.

Our interviewee from the school said:

I do data entry at 5 AM because the system overloads later in the day [.the system begins to buffer[

Interoperability of the state level database, time taken, persons required for data entry and the fear of punishment also pose challenges to data collection. Each state and each database also poses its own set of challenges. Two schools based in **Uttar Pradesh and Uttarakhand noted the difficulties in UDISE + data entry of children** who are transferred from other schools. **S4 said,**

"We couldn't enroll a lot of students even if they were studying at our school, since it showed that they were already enrolled in another school, there are all kinds of hassles"
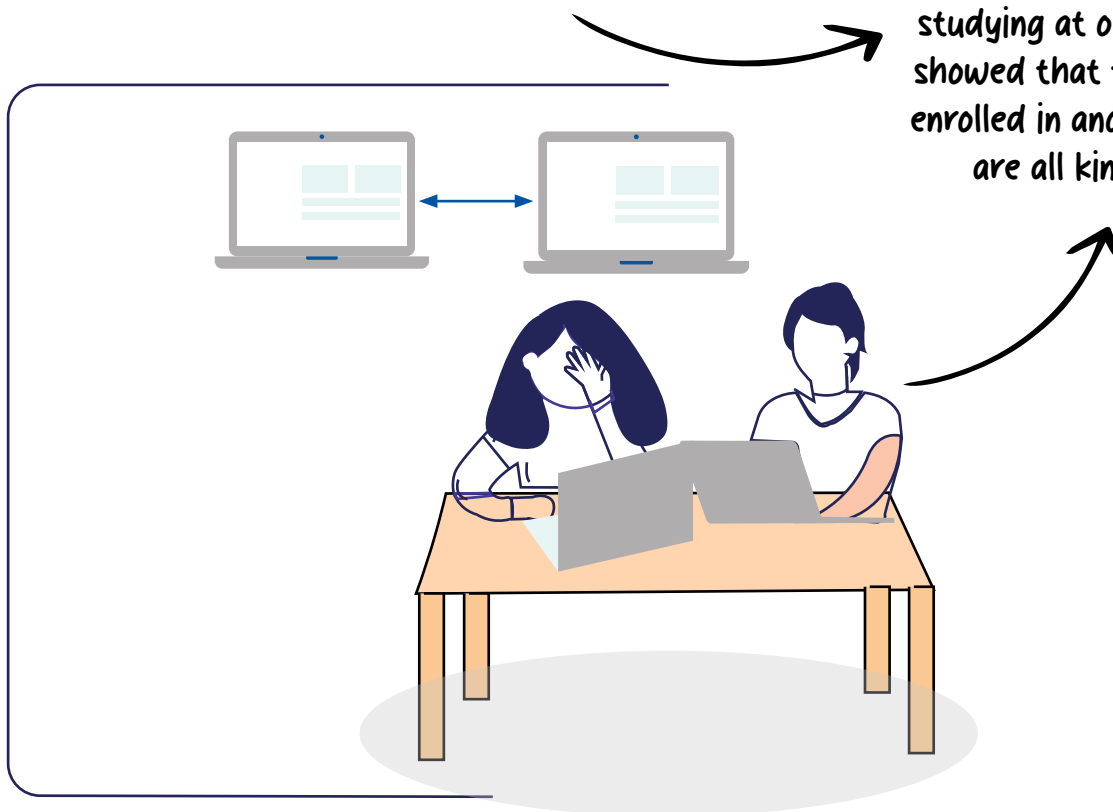
2. Upstream Challenges

Schools face challenges in collecting data and uploading it on to the requisite databases, as they have several dependencies on the government. They also do not get timely resolutions to their queries to meet the deadlines for updating the data.

2.1. Government support is not always available

S6, a school in Assam reported:

"
People from the State Government are on Whatsapp. So I feel that they are helpful, But I feel there is a lot of confusion happening all the time[our school is well-known, but sometimes, other schools, there is a lot of rude language used[
"

S2, a school in Tamil Nadu reported that:
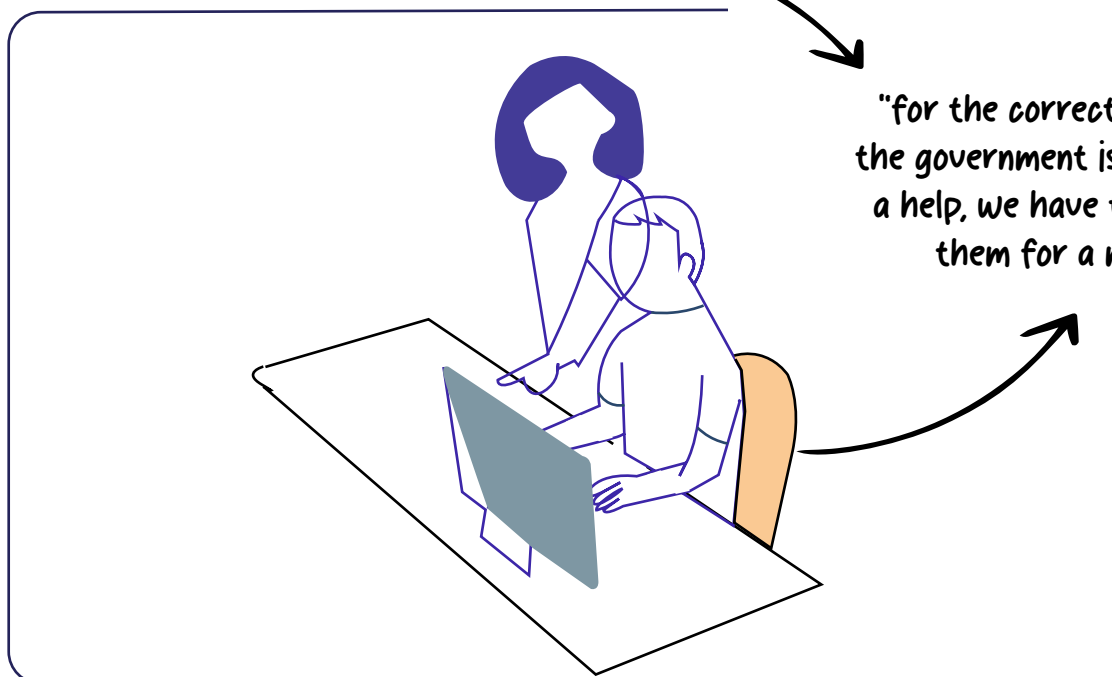
"for the correction of data, the government is not much of a help, we have to be behind them for a month"[

2.2. Data is required at short notice

Two schools and one field practitioner reported that the government contacted schools for additional datasets without giving them sufficient time. **S6 said,**

"Many times what happens is [..they don't give (us) proper deadlines. it is always either two days, one day, so that is very difficult for us to do"

3. Data collection Is a Burden for School Staff

Only three schools and one field practitioner viewed the data collection exercise as an easy process. Most of the field experts we spoke to alluded to data reporting casting a burden on teachers. When we probed the schools, we discovered that the process is time consuming, burdensome and viewed as a compliance measure. Although, **S1 stated:**

> Once a person gets familiar, it's easy to navigate[It is a habit[Most of the information they ask is what we collect anyway

A field practitioner based in Karnataka reported,

> When they introduced it initially, (teachers) felt it is difficult 80-90% are now telling it is very helpful for us compared to previously, burden is now lesser, this is the kind of feedback we are getting.

S1 also faced difficulties in collecting Aadhaar data of students under the revised mandate of UDISE+. **They said:**

> But from last year they have been asking for individual student data, including their Aadhar. It becomes difficult for us, because their names might not match. Sometimes they are already enrolled in a primary school and from there their name has not been struck off. In UDISE, we cannot do anything because it shows that the child is already in another school. Most likely though, this will get streamlined this year.

But **another field practitioner based in Karnataka pointed out that:**

"Teachers filling this are not really people who want to be doing this. Teachers are made to do more than what they're paid for...teachers were doing data entry at 11 12 in the night due to (application) capacity issues"
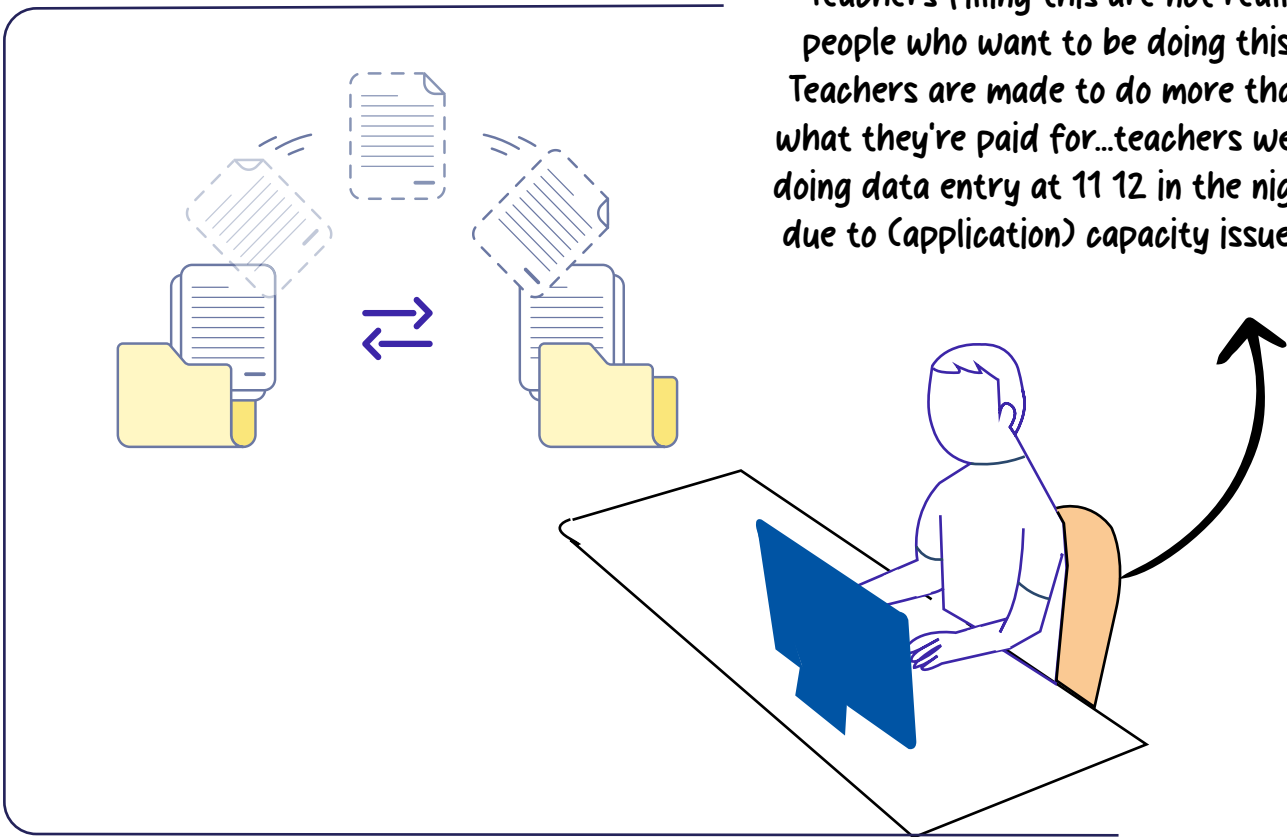
A policy expert based in Delhi also reflected ,

Everyone sees it as a burden, they are not "intimately connected" to the exercise

He went on to highlight how the data caters to an interest group that is far removed from the group that actually collects and reports the data. **He called the former, "data bugs" and the later "data victims:"**

Data is the concern of data bugs, everyone else is a data victim who have to carry the burden like climbing a mountain

> **Research Team insights on the linear flow of data in the school data ecosystem**
>
> Though we conducted a small number of research interviews, we identified a few recurring creases that required to be ironed out in the process of collecting and reporting data to state mandated databases. Firstly, technical challenges encompassing technology infrastructure , data requisitioning process and training gaps had a strong influence on the data collection experience. Secondly we inferred that the "upstream" flow of data has resulted in a power dynamic that has created a group of data masters (of mandatory reporting systems such as UDISE+) who wield control over data victims (school staff who collect data). Data victims don't have agency over (the hows and whys of) the data mobilized by them, and resultantly erect walls between themselves and the data, creating a dissonance. This insight is further bolstered by findings detailed in the following section on Barriers to Data Utilization.
>
> A research study with a larger sample set will help to contour the role of data in the corridors of power and the impacts on the quality and utility of data. Ethnographic studies would help to bring more nuanced observations surrounding data practices, resources and experiences.

C. Barriers to Data Utilisation

Although UDISE+ data was being collected from all recognised schools in the country, the use of the data collected is restricted to certain stakeholders and certain uses within the education ecosystem, leading one to ask the question - what are barriers to the use of education data? **In the words of P2**,

> ❝
> I have not seen it (a data culture) in the Indian system yet. As I said, many states are now only starting to move from compliance to data use. Data use is still nascent in many states, including even Tamil Nadu.
> ❞

He further added insights on the situation at the school level:

> ❝
> The teacher will work with their intuition inside the classroom. So I think unless that culture of consuming information and data and making small decisions based on that is built, I doubt if very much will be used
> ❞

The culture of data-driven decisions has not permeated down to the level of schools. The interviews allowed us a brief insight into why schools may not use data (where available) or demand for utilitarian data (where not available).

1. Limitations of the data collected

1.1. Data collected does not offer customized insights and only serves the macro level of the education ecosystem

Three field practitioners and two schools observed that the data collected only gives them overall insights into the ecosystem without providing insights at the school level. P8 stated that the insights from datasets lack school-level configurability. **He emphasized:**

"UDISE does not give the teacher the data cuts that they need–the teachers don't get the (necessary) data cuts at all"
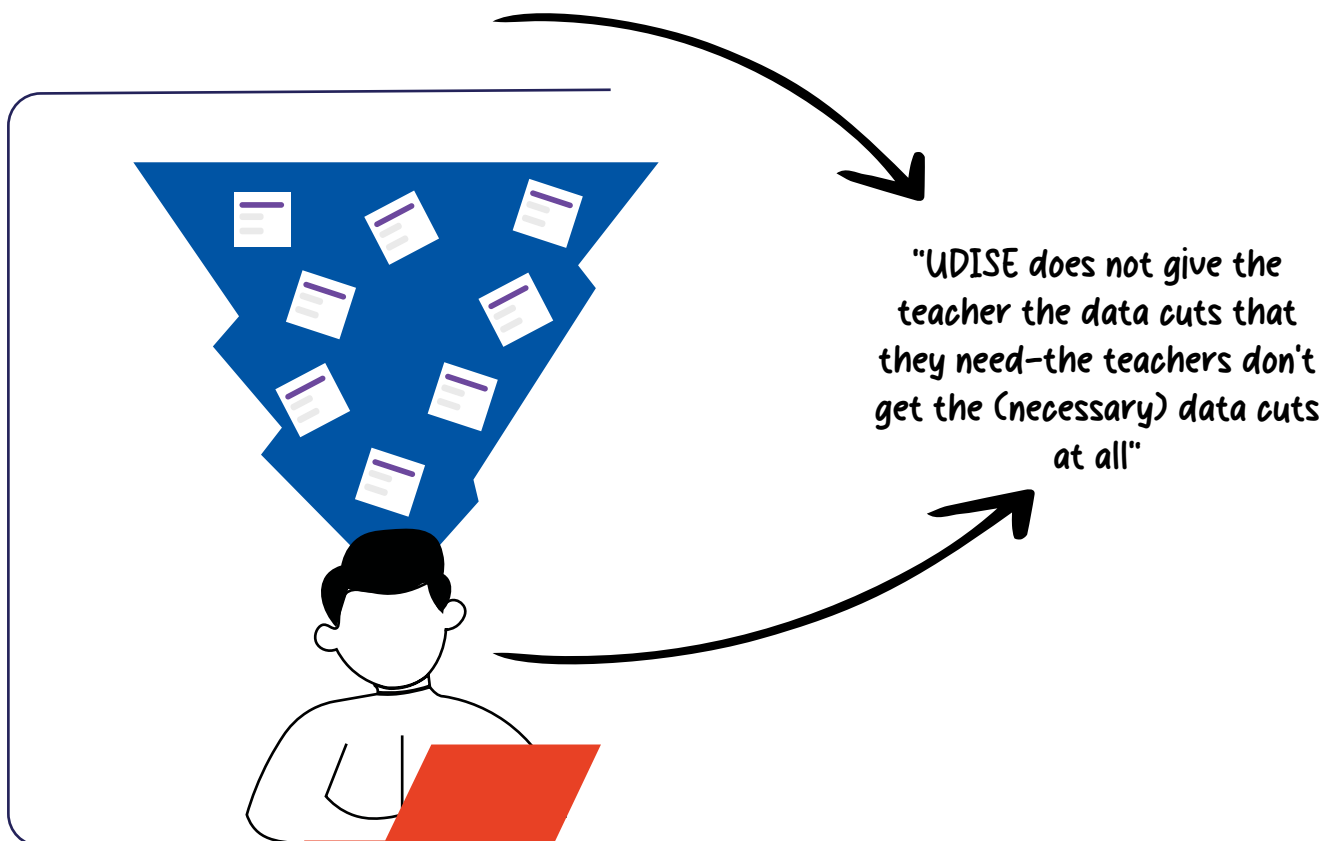
P2 gave an illustration of this limitation:

> UDISE data is useful for high level trends... As I said, pupil-teacher ratio, GER, net enrollment, number of schools, right? It's good for descriptive statistics. It gives me enough data to, let's say, chart a distribution curve of what the PTR (pupil teacher ratio) looks like district wise etc...While UDISE will track the number of teachers trained, UDISE is not tracking what is the quality of that training delivery. How many teachers, let's say, mastered whatever they were trained on? So there is nothing that even remotely indicates education quality metric.

S3 noted that while benchmarking insights are provided by the government, this is given on an overall basis, and more granular data would be useful at a school level:

> We would like to know how our school is performing in comparison with government schools as our students are primarily from the neighbouring slums, there is no point in comparing ourselves to a fancy private school.

P5 said,

> UDISE does not cater to spaces which are unique[. infrastructures of safety (including emotional)are non-negotiable especially to where we work

Customised data insights is thus the biggest challenge to data utilisation. However as seen in Table A, the insights from mandatorily reportable datasets such as UDISE+ mainly informs decisions made by the government and civil society initiatives that work to influence policy at a high level, The data does not provide utilitarian or actionable insights at the school level. **P2 validated this assertion with his observation:**

> ❝ I think that is one of the primary challenges with UDISE – while it is excellent to get the data that the Ministry is looking for, it doesn't really allow for the degree of customisation and the speed of customisation that the specific states may have. ❞

1.2. Data is inaccurate/inconsistent

UDISE+ data, though widely used, has been criticised for its inaccuracy. The lack of accountability for updating incorrect information,[14] coupled with the schools' fear of reprimand against non-compliance, have resulted in the reporting of inaccurate data.
S1 observed:

"Fake or real, for RTE purposes, all schools are compliant"

---

14 Mridusmita Bordoloi and Varun Kapoor, 'India: Using open school data to improve transparency and accountability' (UNESCO) 2018 available at <https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000265933&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_43410c09-152e-469c-aac5-0d7ab4289f9f%3F_%3D265933eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000265933/PDF/265933eng.pdf#%5B%7B%22num%22%3A192%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2Cnull%2Cnull%2C0%5D> accessed 2 November 2023

Aadhaar data entry and verification also a challenge for schools, leading to discrepancies in the data collection and reporting. **S3 reasoned that this was why**:

> **"** the student's status remain(ed) unverified on SATS till the parent got the discrepancy corrected with the government authorities. **"**

P9 also noted that inaccurate data is a complex problem. He acknowledged that practices such as inflating the number of students for the midday meal does exist but the lack of Aadhaar card based verification in Assam (until recently) posed the issue of duplication of entries.

P6 told us that inaccurate or inconsistent data can be caused by multiple factors. She said:

> **"** Data manipulation by schools may be because of corruption, I have seen with my own eyes, teachers saying to the person who is in charge of bringing ration, bring some ration to my home also.... But there are two sides of the same coin: some children don't have Aadhaar, their parents don't care, so teachers manipulate the number of students in higher classes so they can feed the children. When I go to the school, I ask how many children are offline, 53 children were offline, only 40 were uploaded on UDISE. **"**

Data inaccuracy is a huge barrier to data utilisation. P2 elaborated:

> **"** If the teacher is reporting enrollment as 10 or but the enrollment is five, there is no real way for us to know how many such instances are happening. **"**

Three schools also pointed out the data is of no use due to inaccuracy. The observations of S2 alludes to this challenge:

> ❝ No no we cannot use EMIS data for our administrative purpose [– this data is not 100% ❞

While stating that the UDISE+ data was of no use to them, S5 said:

> ❝ You are judging a school based on the information it has provided by itself. This may be fudged. Then what is the point? ❞

1.3. Real-time data is not available

The uploading of UDISE+ data was made online to ensure real time exchange of information.[15] Two field practitioners and a school in Assam (S6) spoke about how the data is not updated often in Jammu and Kashmir and Assam. However, this may be a state specific problem. **P5, the practitioner from Kashmir stated,**

> ❝ So, one impact of conflict is loss of rigour. And data requires a lot of rigour. I feel I keep saying this but that loss of rigour is because there is so much uncertainty and disruption that even if you wake up with the motivation to apply that data there are things out of your control that, doesn't let it happen. ❞

P9 noted :

> ❝ The data is there but it's not very up to date. There are some what you call not processing time, I would say data updation time. What I have understood is that in 2021, the last data was updated. ❞

---

[15] 'UDISE_Booklet.Pdf' <https://dsel.education.gov.in/sites/default/files/update/UDISE_Booklet.pdf> accessed 23 September 2023.

Data also poses a problem in that it is constantly changing. **S2 noted that:**

"We collect information only on a yearly basis, if parent's shift residence mid-term, it becomes a problem"

1.4. Quantitative, rather than qualitative datasets are prioritised

The focus of UDISE+ is primarily on quantitative indices, such as enrolment rates, drop out rates, school infrastructure, teacher qualifications etc. The UDISE+ data informs administrative decisions such as allocation of budgets. UDISE+ does not address student learning outcomes or other qualitative indicators of student well-being. This was pointed out as a limitation of the dataset by some of our interviewees. In the words of p2

While UDISE will track the number of teachers trained, UDISE is not tracking what is the quality of

There is nothing that even remotely indicates education quality metric.

P6 said

"There is a focus on quantitative over qualitative, quantitative does not give the entire picture"

This points towards a flaw in the design of databases such as the UDISE+. **P2 pointed out how UDISE+ was never meant to solve the 'quality problem':**

" If I'm not wrong, the purpose of the UDISE itself is to make sure that the leakages in the fund disbursement and utilization reduces over time. I don't think it is meant to solve the quality problem, if I'm not wrong. But obviously these things evolve over time. Where UDISE started and where it is now, it's quite different. They have started looking at a little bit more of school, infra metrics etc, that they have borrowed from the NIEPA or the SECI indicator list as well. "

Two other field practitioners in Haryana also reasoned that the focus is on collecting quantitative data since evaluating qualitative data is objectively harder.

P2 described the difficulty in capturing the complexities of the education system in the dataset.

> UDISE+ data is used to track India's progress towards Sustainable Development Goal 4 **["SDG"]**, which in itself relies heavily on quantitative measures to **"ensure inclusive and equitable quality education and promote lifelong learning opportunities for all".** This may also be why the focus has continued to remain on quantitative indices. The specific targets outlined under SDG 4 encompass various aspects of education, from ensuring access to quality primary and secondary education to promoting lifelong learning opportunities for all.[16] Barring outcome related measures, all other indicators are tracked using **UDISE+ data**. This includes data on enrolment rates, school infrastructure, teacher qualifications etc.

> "Typically, what defines quality of learning is what happens in a classroom on a day-to-day basis. Yes infra, yes, all of that yes, but at the end of the day, it comes to how the teacher is teaching...Only if the trainer is good, will the teacher get trained well. Only if the teacher is trained well, there is hope that the teacher will teach well, and only if the teacher will teach well, there is hope that the student will learn and achieve whatever the end outcome the reform is towards."

This was echoed by the practitioners. Five practitioners and one school spoke about the importance of learning outcome data and how it is overlooked by UDISE+. **P8 said:**

> "Government generates broad data. We cannot understand the challenges in learning (outcomes) from this"

---

[16] National Statistics Office, Sustainable Development Goals National Indicator Framework Progress Report 2023 available at <https://sdgindiaindex.niti.gov.in/ner-assets/Files/Metadata_NER_SDG_Index.pdf> accessed 24 November 2023

P2 also pointed to the focus on a few metrics as a limitation of the UDISE+ dataset. Driving focus to other metrics would be a difficult process.

> "
> The next time you want to do it for a different metric, you will have to again publicize it. The intention to work and the culture to work will be there, but the metrics will have to be, changed over time. And it will always have to be sharp. You can't give 35 things and expect all of that to happen. However, agreeing on such sharp one or two metrics, by the entire education fraternity will be a challenge.
> "

---

[17] Government of India Department of School Education and Literacy, Report on UDISE + 2020-21 available at <https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/UDISE%2B2020_21_Booklet.pdf> accessed 24 November 2023

2. Lack of trust among various stakeholders in the education data ecosystem

DISE was initially introduced to plan and monitor the District Primary Education Programme launched in the early 1990s.[17] Over the years, DISE and UDISE+ data has evolved to be the foremost government source on almost all indicators relating to schools. While incremental improvements have been made to benefit schools on the data collection and reporting systems, such as a transition to an online model, UDISE+ primarily continues to serve the Government and government decision making. Resultantly, the schools **("data-bugs")** view the data ecosystem with distrust. P8 narrated his experience in approaching schools on data:

> The first time we collected the data, (from schools), schools were afraid

He added that the fear was overcome only over time, after the schools realised that the practitioner's organisation used the data to design effective classroom practices. Hence, even the practitioner had to intentionally and carefully build the trust of the schools to garner quality and actionable data.

2.1. Schools fear reprimand based on their data inputs

This lack of trust in the reporting system seems rooted in a fundamental scepticism, driven by fears of facing repercussions for failing to upload accurate data or based on past negative experiences that schools may have encountered. This distrust may be fueled by concerns about consequences, leading to a cautious approach. **P8 elaborates on how schools react to data,**

> They (schools) don't see themselves as users of data, they see themselves as victims of data.

S4 also felt similarly about the data, **our interviewee said:**

> They (DEO/BEO)also use the data to kind of give the direction to the State that which school should be shut and which school should be

P8 narrated an incident where despite a reduction in the number of indicators to be collected aimed to reduce burden, people still continued to collect all the 85 items. **They reasoned:**

Our interviews revealed that schools often viewed data reporting mandates as whips that would be leashed upon them. P4 **stated that :**

"Suppose they suddenly ask for it?"

"The mindset towards data has been crafted in a way that data is your mirror of how you perform. If there is any sort of a letter notice, it is taken as compliance and there is a lot of fear attached."
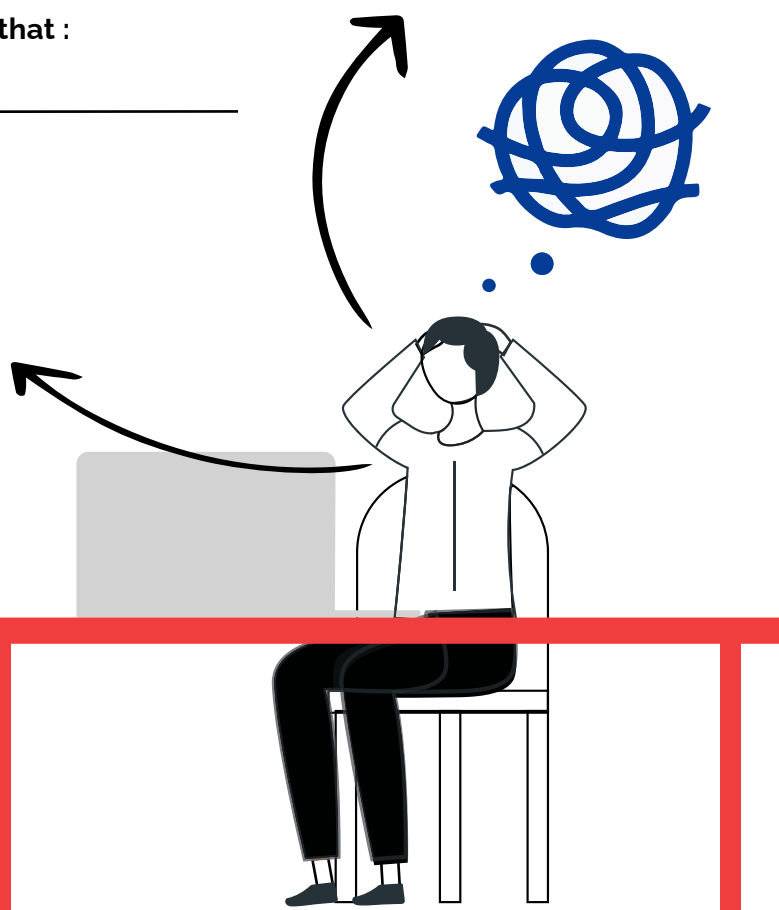
2.2. The practice of data use has not penetrated to the school level

At different levels, diverse stakeholders in India have increasingly embraced the use of data to inform decision-making processes. However, practitioners are still sceptical of the level of data use and data driven decision making in India. P1 and P8 observed that the practice of evidence based policy making is underutilised in India. P8 shared:

> There is no evidence that evidence works[where it is written is more important than what is written.

According to P2, data use has in fact picked up in the country, but it has not traversed downstream from the district level.

> Data use has penetrated in most districts. They haven't penetrated beyond district level, or I would say further down the value chain.

His views on the same were corroborated by two other field practitioners. However, a group of field practitioners based in Karnataka shared their experience on the presence of data culture at lower levels too. They worked closely with the BRCs and CRPs in Karnataka and shared examples of data usage by these officials.

> School infra data is being uploaded on the database by schools if some company or CSR is working in the block, the block officials will go through UDISE+ plus, it will help assess whether there is a requirement for infrastructure in any of the schools in the block,, this is an existing practice. They (BRCs) always motivate HMs to go for alternative budgets like taking CSR funds, politicians support.

However, we did not find this extent of data use at the school level. The reasons for low engagement with data could possibly be the low penetration of internet, low technical capacity, the nature of data insights coupled with an inherent fear of reprimand for showing data that reflected that a school was non-compliant/ not doing well on certain parameters. Three schools felt UDISE+ data was not necessary. The opinion of S4 captured how they feel about the data,

these data insights are not generated for the school's benefit.

P2 said,

> There is a massive gap in terms of the government's capacity to manage such databases and technological tools

> No, We don't need any of this data, right? We don't even for our internal purpose. We don't even need anything to do with caste data, or anything. We just need our academic records and all those things, right. So apart from that for our internal purpose we do not need any of this.

3. Capacity constraints in the use of the Data

The system does not have the capacity to effectively collect the data required. The databases and teachers are not fully equipped for the data collection exercise which adversely affects data utilisation. The government is also unable to effectively leverage the data infrastructure and effectively use the data collected. A lot of data insights that schools told us would be helpful can be generated using the existing datasets that the government is already collecting. And yet,

Further another policy practitioner notes that,

"Whatever is absent, it is absent because they may not know what to do with that data. It can't be put in practice at this point, because the system in itself may not have capacity."

System Error !
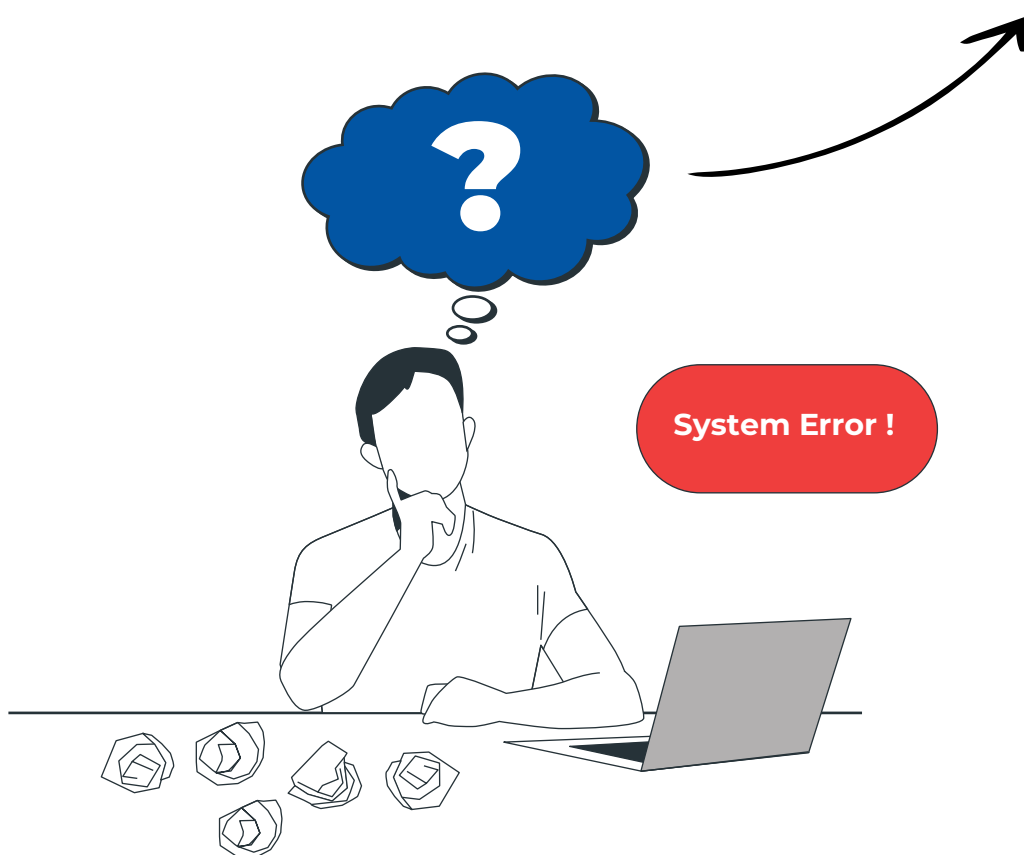
P4 also explained to us the capacity constraints of schools and teachers. He narrated his experiences as a Teach for India Fellow to demonstrate how knowledge of issues does not automatically translate to action at the micro-level.

He goes on to add that not every teacher is in the position to do so. The strain on teachers is more than that of a corporate person or a farmer. The job requires expending both their mental and physical energy. He says that because it is a physically demanding job, it further limits the capacity of the teacher to do anything about the poor outcomes even if they really want to address and improve the outcomes.

> **"** If the National Achievement Survey says Science and Math learning outcomes in Pune are lower than the national average, then it can be advocated that we need to work on Science and Math education. Everyone including the school and its teachers is comfortable with this macro level data and it is accepted. They are also accepting and welcoming of other interventions and NGOs that provide this support. But at a micro level, if you take my example as a Teach For India Fellow, I had a classroom with students in three distinct groups of learning levels. So that means that in the same classroom, with the same students, I am teaching in three different methodologies. **"**

4. Schools are unaware of the potential uses of the data collected through government mandates

S6 spoke about how the UDISE+ data does not show any additional data, beyond what the schools already have:

> ❝
> It is not showing us any new ideas, it is our data only.
> ❞

Considering the immense potential that can be unlocked by a database as comprehensive as UDISE+, the schools were asked what data insights, if provided by the Government, would be of utility to them. One practitioner and one school each mentioned that they never thought of it at all.

P9 stated,

> ❝
> Honestly, I never thought UDISE could be used in such a way. But after I received an email from (the Research Team), I just tried to do some research on how UDISE can be used. And I see a lot of potential.
> ❞

Potential Uses of Data (Collected for Reporting Mandates) by Schools

Five schools were able to give us insights on data that they would find useful. Besides the insights gained from the interviews, the research team also analysed the Microsoft Excel sheet generated by UDISE+ to come up with potential insights that the data may provide. The findings from the interviews and our analysis has been synthesised below.

1. Learning outcomes: Five field practitioners mentioned that they would find data on learning outcomes helpful. However, only one school mentioned it as a useful insight. Understanding learning outcomes is important to achieve the goals of educational policy, and to hold accountability.[18] P8 also emphasised the need for focus on learning outcomes as India currently does not have adequate pedagogical models to cater to the diversity of students. Currently, learning outcomes at the national, state and district levels are determined by three surveys: ASER, NAS and the India Human Development Survey. However, all of these surveys gather data through sampling and do not collect data at an individual level- UDISE remains the only database which does this. Incorporating learning outcomes data in the UDSIE database would enable schools to undertake time series analysis to observe changes in learning outcomes. Such an exercise would be useful to assess the loss/ gain of learning over the years. For instance, it would have been useful in ascertaining the loss of learning due to COVID-19.

2. Convergence of datasets: Three field practitioners mentioned the importance of converging UDISE+ and other such educational databases such as socio-economic and health databases. P2 mentioned how his CSO had linked the stated EMIS with the NRHM database to converge the PICME number of the student with the student's unique ID number. However, P2 also pointed out the wide gap between high-level conversations on converging such databases and the missing action on the ground.

> " When you talk about the Ministry level, we talk a lot about convergence between the Social Welfare Department, Education Department, Health Department. But when you go to the ground convergence, it's missing. "

---

[18] Johnson D and Parrado A, "Assessing the Assessments: Taking Stock of Learning Outcomes Data in India" (2021) 84 International Journal of Educational Development 102409 <https://doi.org/10.1016/j.ijedudev.2021.102409>

3. Demographic data: Two schools elaborated on the need for more granular demographic data. S4 explained the use of socio-economic data for school purposes:

P5 emphasized the need for demographic data to understand the lived experiences of the students they work with:

> For us, we're a CBSE school, I can go check the facilities and infrastructure of other schools with CBSE. So for us, the kind of data that would be useful has more to do with the economic status of the children; if they have BPL cards, then they can be eligible for some kind of scholarship. So those are the things that we would require

> I want to know how many of my children belong to historically displaced communities in the area; how many of my children come from, you know, the majority community within Kashmir itself. There's also class based segregation in Kashmir, just like other places, so I, I feel like that's definitely something- just to be able to view every context uniquely for the uniqueness that it brings and then add those elements to the data metrics tracked

4. Benchmarking: Three schools mentioned how they could benefit from benchmarked insights from the Government. While two schools had clear thoughts on these, the third school acknowledged the need for such insights only when prodded. Currently, the UDISE+ database provides benchmarked insights, but not at granular levels.

> *We would like to know how our school is doing in comparison with other government schools, since we have students from the neighbouring slums, it would make more sense to compare to government schools to see what impact we are creating. The current insights give the school its performance in comparison to all schools in Karnataka. The bigger picture of data insights would be helpful to know data insights such as enrollment ratios in the state, learning outcomes so we can measure our schools performance against other schools.*

S6, another NGO run school, based in Assam, spoke about benchmarking needs beyond academic indicators:

> *Data insights, that would be helpful, at least some benchmarking, a small example of the number of books per school. We want to be a model school, we want to see data in that way. Time table wise, how much time schools are giving per subject, sports, arts, how time is allocated, it would be interesting to see how much importance is being given to extra-curricular activities*

5. TC portal: One school mentioned that they would find a system that enabled seamless transfer of student data between schools useful.

6. Non- academic data of students: Student data that encompasses more than just academic achievement metrics, such as critical thinking measures. Currently, UDISE+ collects data on Children with Special Needs ["CWSN"]. However, correlating this with the number of teachers who have been trained to educate CWSN, early identification, and the percentage of teachers who use special learning resources for CWSN etc can be helpful in highlighting the training needs for teachers equipped to deal with such children.

Other data: Policy practitioners mentioned that they would find more data on teachers, as well as student data disaggregated by gender to be useful. Schools brought up data insights on attendance rates, attendance of parents in meetings, books read by children, curriculum, textbook lesson plans and assembly attendance related data as insights that would be of use to them. Besides these, the Research Team has also identified the impact of medium of instruction on the performance of the student, attrition rates of teachers, a comparison of training requirements for teachers vis-a-vis training received, as insights that would be of use to schools.

S6 pointed out that the reasons for discomfort towards data among schools and teachers is because they are not made aware of the utility of the data collection exercise, in the absence of the same it becomes only a "burden".

P10 also said that schools he worked with want to know why data is being collected.

Research Team Insights on Incentivizing Data (collection and utilization) for school

The small number of interviews provided a sneak-peek into the reasons why the schools do not use the data (reported on the mandatory reporting databases) to inform their decisions. Responses ranged from the lack of capacity to respond to the data-insights to the problem of lack-of school relevant data. Poignantly, issues of distrust of the data and distrust in the system recurred across conversations with schools and field practitioners, leading us to conclude that the data in the education ecosystem has taken a paternalistic connotation, that has resulted in schools serving the system, but the system not serving the school. More detailed research inquiring into what data insights a school might be interested in, and importantly how to make this available to schools without blaming school teachers for poor data insights, is critical. This would entail a participative approach to formulating databases to which schools are expected to report mandatorily.

Additionally, the value proposition presented by **"data as a service"** to the education ecosystem seems to be on the rise, as schools begin to embrace the data-informed culture. Data as a Service is also an excellent value proposition for NGOs and civil society initiatives that are increasingly using data to determine their programmatic interventions, fundraising strategies and in their policy advocacy activities. Social purpose companies offering Data as a Service may then see a proliferation in the short term.

Part IV. Recommendations

The various challenges encountered in the collection of data for reporting to mandatory databases at the central level (e.g. UDISE+) and state level (e.g. SATS) indicates an imbalanced, top-down power dynamic driving data collection and reporting by data workers. This power dynamic creates resentment about data work amongst data workers, and also affects the quality (and thus reliability and utility) of the data. An obvious outcome created when data becomes available is the question of - who is accountable for points of concerns that the data brings forth? Schools and data workers (teachers) have felt scapegoated and penalised for narratives of high student attrition, low attendance and poor learning outcomes of students, when in fact student learning outcomes are attributable to systemic and not singular factors. Though technology-enabled data collection and reporting has become widespread, this has not necessarily translated into ease of collecting and reporting for data workers, elimination of parallel or repetitive data collection activities for administrative bodies and a full understanding of the motivations for collection of education data (transparency of motives) among stakeholders across the data supply chain. Importantly, this exploratory research shows evidence of inequitable data utility and accessibility outcomes for data workers (schools and school teachers who collect and report education data) compared to the administrative component who are higher placed in the data supply chain. This limitation in utility and access to data has resulted in the practise of data informed decisions being restricted (at best) to policy making corridors or funding decisions for civil society initiatives in the education ecosystem, and has not permeated to use at the school and classroom level.

In respect of data utility, our research did not create evidence of causality between lack of trust in the data affecting the propensity to utilise the data. But the exploratory study established early signs of (potential) correlation between distrust in the data collected and utility of the data at the school level. Limitations in the utility of the data is further compounded by the technical skills to conduct data analysis,lack of tools/responses to the insights from the data and lack of awareness of the benefits of utilising data at the classroom level.

These experiences of data workers in the data collection and reporting process of education data in India calls for a systematic review and reconsideration of the data collection design and process for mandatory education databases, particularly viewed from the lens of the data worker. This will aid towards democratising the entire data supply chain and life cycle. We propose the following prompts to reflect on the current public education data collection process and towards strengthening the value of data for all stakeholders in the data life cycle.

Data Accessibility and Utility Framework

With increased technology driven data-driven approaches by the government in respect of education and in policy making, such as the APAAR ID,[19] there is a need to inculcate equitable data collection habits, that discern data collection practices in the private sector from those employed in the private sector, to ensure equitable development outcomes and minimise harm from data. We propose the Data Accessibility and Utility (DAU) Frameworks as two separate matrices of principles and indicators that influence education data utility and accessibility at the level of the school unit to introduce equitable utility and accessibility to the data collected and reported under mandatorily reportable education databases. Our exploratory research has shown these principles and factors to have an influence on equitable data utility and accessibility outcomes, but further research is needed to determine the extent of weightage each component has on the overall accessibility and utility of datasets.

Scope for Further Research

1. How may a single public education data collection system serve several stakeholders equitably in the education data supply chain?
2. Are current public education data collection systems built to serve the data needs of schools and classrooms
3. What public education data needs to flow upstream, downstream and what data need not? Who decides?
4. What can help to build healthy accountability mechanisms for education ecosystem stakeholders in respect of narratives created by the data?
5. What are ways to drive accountability other than penalty based methods?
6. Across public education data ecosystems, what kind initiatives may help strengthen accuracy of data?

[19] Akhil Bhartiya Shiksha Samagam, " Thematic Session on National Credit Framework (NCRF) & Automated Permanent Academic Account Registry (APAAR)"(Government of India) 2023 available at https://www.education.gov.in/sites/upload_files/mhrd/files/nep/ABSS_Report_Session_3.pdf accessed 24 November 2023

Data Accessibility Indicators:

| Principles	Indicators		
The system is able to overcome technological barriers	The data can be entered without an internet connection	The system does not crash	The data can be backed up and retrieved even in case of system crashes
The system is easy for the users to understand and use	The database comes with a data dictionary which explains the technical words used in the database	The government provides necessary human/other resources for additional support	The time taken to enter data is less
	It is possible to auto-fill fill data	The database is aligned with Web Content Accessibility Guidelines	The database should enable import of electronically stored data
	The data can be revised without difficulty	The database is interoperable with other EMIS	The language of the database can be changed
The database has adequate safeguards to ensure security and privacy of data.	The database can be accessed only upon logging in using a secure password	The database has standardised privacy and security protocols	The database asks for consent from users before sharing information with other EMIS
The Government provides support and training to the users.	The Government provides both hardware and software infrastructure	Technical support from the Government is available for the users	The reporting is done only through the database, avoiding out-of-system reporting

Data Utility Indicators:

Principles	Indicators		
The data is accurate	The database has a provision for vetting the data by school/government/third party	The data is up to date	
The system is transparent	The data reporting requirements will be intimated to users through a pre-annouced schedule	The users are informed of the implication if they do not fulfill the reporting requirements within the specified timelines	The users shall be notified of what happens to the data that flows upstream and the purpose of the same
The data insights are easy to analyse	The data has been provided in analysable formats (excel sheets/databoards)	The data does not have too much technical jargon	The data can be used to inform decisions at school level
Data insights pertinent/actionable insights	Sub-aggregates of the data are available on request	Useful and relevant datasets (data comparisons across time, geographics based on board affiliations etc)	Convergence with other sectoral databases (health, nutrition, etc)
Miscellaneous	Schools can seek entitlements based on the data	The data helps users identify the status of their compliance with government mandates	The databases is compatible across different devices
The use of data does not conflict with school	The Government should provide constructive feedback and not reprimand the schools	The datasets should protect student anonymity	The school should have access to grievance redressal mechanisms

Other Recommendations

1. Build trust in data collected and data collection processes for all stakeholders (student, guardian, school teachers, schools, education functionaries, policy makers, and funders.
2. Build awareness of significance and utility of data at the unit level (schools) as a muscle to nudge unit level inclination to use data and make data informed decisions.
3. Build unit level (school/ school teacher) ability to use data meaningfully through varying levels of data analytics capabilities.
4. Build processes to verify that the system is working for all by design and not merely to fulfil a top-down and cosmetic need for data.
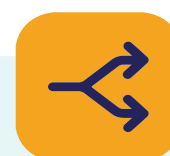
The Way Forward in Data Empowerment of Schools

This research study has only scratched the surface to open up narratives of data workers in education data collection, data workers' lived experiences pertaining to data collection and data workers' perception of implications of the data they collect and report. It has also created evidence of inequitable data utility and accessibility for participants such as schools and school teachers (data workers) low in the data supply chain, than for participants such as public administrative instrumentalities that are higher placed in the data supply chain. We have created a formal acknowledgement of the challenges in collecting and reporting data at population scale, and the potential utility implications of non-democratic data related processes. Insights from this exploratory study would be useful to inform further research at the intersection of equitable data collection practices, utility and accessibility implications of data collected through such practices. Further directions of recommended research are as follows:

1. What are the field-level challenges that teachers are facing in collecting and reporting data to compulsory education databases (such as UDISE+/ other student assessment databases)? Does this vary by demography and type of school?
2. What are perceptions at the school level in respect of utilising data to make data driven decisions in the classroom? Does this vary by demography and type of school?
3. What are the constraints of schools to utilise the data in their classroom (time, knowledge, resource capacity)?
4. How is data from compulsory education databases used to develop AI systems? are the use cases of data (in the time of AI)
5. What type of consent related practices exist in the collection and reporting to compulsory education databases?
6. How have compulsory data collection and reporting databases influenced decisions at the policy level?

ANNEXURE

ANNEXURE I- Methods Used to Ascertain Gap in Literature

Purpose

Keyword mapping of literature to find the research gap to establish the importance of the Report objectives

Approach

Based on the inputs, VOS Viewer was used to undertake the task. It is a software tool for creating visualisations referred to as maps based on network data for visualising and exploring these maps. The database of the research papers was compiled using Lens website. The keywords that are related to the study such as "Education data management system", "UDISE+", etc were used to identify the relevant papers. The collection of papers is used as input for VOS Viewer to build a graph based on the text in the title and abstract of these papers.The keywords selected had appeared for a minimum of 5 times across the papers.

How to read the keyword map?

In the map below the keywords are referred to as nodes. Nodes with specific colors form a cluster like blue, red, yellow, green. The clusters include a set of closely related nodes. The size of the circle implies the weight of each term based on co-occurrences i.e nodes like "data", "education", "information" . The strength of relationship between two nodes is defined by the thickness of the link connecting them and the distance between two nodes.

A. When the focus is on the node "Data"

In Figure 1 (in B6), "data" and "education" are part of different clusters. But there is still a reasonable connection found between these two keywords across research papers as shown by the close proximity and thick links. However, there is a weak relationship between "data" and the "education sector", demonstrated by the distance between them. It implies a lack of research in the education sector in the context of use of data.

Secondly, the node data is closely connected to nodes in the same cluster like "DISE", "data collection", "elementary education" etc. but "data" is not close to "teacher", "school management", etc revealing limited research in these intersections in India.

B. When the focus is on the node

Figure 2 in B6 shows that the node "UDISE" is closely related to "school", "database","management information system", etc. However, across the studies there is limited association in prior literature between "UDISE" and "decision making", "outcome", "analytics" etc in India.

C. When the focus is on the education sector

Literature points towards a disjointed relationship between the education sector and data as they belong to different clusters. It implies that the role of data in the education sector is not well established in the existing literature. Similarly, the nodes, "education sector" and "EMIS" (Education Management Information System) are distant from each other. However, the relationship of "education sector", "decision making", "big data" shows the emerging relevance of such concepts in the education sector. This has been mapped in Figure 3.

ANNEXURE II- Consent Form Circulated in a Selected School

CLASS 11   Student data form 2023-24

UDISE DATA COLLECTION

anagha997@gmail.com Switch account

* Indicates required question

Email *

☐ Record anagha997@gmail.com as the email to be included with my response

DIVISION *

○ A

○ B

○ C

○ D

ROLLNO *

Your answer

NAME OF STUDENT IN BLOCK LETTERS *

Your answer

GENDER[1- MALE, 2 -FEMALE, 3 - Transgender] *

○ 1

○ 2

○ 3

DATE OF BIRTH[DD-MM-YYY] *

MM DD YYYY

___ / ___ / ___

MOTHER'S NAME *

Your answer

FATHER'S NAME *

Your answer

GUARDIAN NAME (if any)

Your answer

STUDENT AADHAR NUMBER *

Your answer

STUDENT NAME AS PER AADHAR CARD *

Your answer

PRESENT ADDRESS *

Your answer

PINCODE *

Your answer

MOBILE NUMBER OF PARENT *

Your answer

ALTERNATE MOBILE NUMBER

Your answer

E-MAIL ID OF PARENT *

Your answer

MOTHER TONGUE *

Your answer

Show the previous page

SOCIAL CATEGORY[1- GENERAL, 2- SC, 3- ST, 4- OBC] *

○ 1

○ 2

○ 3

○ 4

MINORITY GROUP [1- MUSLIM, 2 - CHRISTIAN, 3- SIKH, 4- BUDDHIST, 5 - *
PARSI, 6- JAIN, 7- NOT APPLICABLE]

○ 1

○ 2

○ 3

○ 4

○ 5

○ 6

○ 7

WHETHER BPL BENEFICIARY[1- YES, 2 - NO] *

○ 1

○ 2

WHETHER AAY(Antyodaya Anna Yojana)[if BPL beneficiary, 1- Yes , 2-No] *

○ 1

○ 2

IS THE CHILD AN INDIAN NATIONAL (1- YES, 2 - No) *

○ 1

○ 2

ADMISSION NUMBER *

Your answer

Submit   Clear form