

The Numbers Game

How Well Has It Served the Cause of Education?

KIRAN BHATTY

India has built a regular school-based decentralised data collection system investing much time and resources. However, all the data has not helped in determining how far the country has progressed on the goal of “Education for All” given the varied and often contradictory evidence. The quality of the data collected is highly suspect, as different sources provide vastly different estimates and the processes of verification and validation, are not in use. It would appear then that despite the fact that the coverage and scope of data collection has increased enormously with many more indicators added, and technology has enabled better management of data, some nagging questions remain about the quality, utility and purpose of the data, with obvious implications for planning and policymaking.

Two decades ago, in the mid-1990s, I was part of a survey of primary education across five north Indian states, known as the Public Report on Basic Education (PROBE) survey.¹ It was one of the first field-based studies of its kind, which explored the quality of government elementary schools using multiple perspectives like that of the teachers, the parents and the students. The study collected primary evidence on school quality from each of the states, and questioned the prevailing wisdom that centred around poverty and demand factors as the main predictors of educational deprivation. In highlighting the abysmal quality of government schools and the “push-out” factors the report shifted the debate away from poverty, busting several myths about parental motivation, income poverty, and the quality of public schooling. While it went on to be a very influential report, as researchers we struggled to ensure its authenticity in the light of existing data and information.

The primary constraint we faced at that time was the difficulty of finding disaggregated secondary data on schools. All that was available by way of educational statistics were the annually published Selected Educational Statistics (SES) and the different rounds of All India Education Surveys (AIES). The SES data was available only at the state level and even then after a lag of three to four years. The AIES, while more detailed, was also available only after a gap of several years. There were no national level household surveys, which collected information on education regularly, and we had to depend on the census and the education rounds of the National Sample Survey (NSS) (both with decennial time frames) and the International Institute

of Population Sciences (IIPS), National Family Health Survey (NFHS-2, 1998–99), for validation and comparability. But, while our efforts to align our own data with existing sources were largely for research purposes, we wondered how policymakers were managing without an annual time series and any of the details one might expect from a comprehensive and dedicated education data set. This was especially puzzling because the planning process and budget allocations followed, and still do, an annual cycle!

Quality, Utility and Purpose

Today, the situation is vastly different in many ways. A lot of energy (and investment) has gone into building a regular school-based decentralised data collection system. This District Information System for Education (DISE) set up after the Sarva Shiksha Abhiyan (SSA) was launched, and now Unified-DISE (U-DISE, 2014–15)² collects data from all schools and provides report cards up to the secondary stage for every state, district and school. It is remarkable that this data is compiled and school report cards prepared and uploaded on the website on an annual basis.

Education data from households is also being collected by panchayats and compiled annually in village education registers (VERS). A few states have supplemented this with data from child tracking surveys (CTSS), which enumerate the population of schoolgoing children. In addition, the Ministry of Human Resource Development (MHRD) commissioned three rounds of household surveys in 2006, 2009 and 2014. The Social and Rural Research Institute–Indian Market Research Bureau (SRI–IMRB) surveys as they are called collect information on children in the age group of six to 13 years that are out of school. Other large household data sets have emerged too, in addition to the NSS and census, such as the National Council of Applied Economic Research’s (NCAER) Indian Human Development Survey (IHDS-I, 2004–05 and IHDS-II, 2010–11), the Annual Status of Education Reports (ASER), since 2005 and now the Socio-Economic Caste

Kiran Bhatti (kiran.bhatti@gmail.com) is senior fellow, Centre for Policy Research, New Delhi.

Census (SECC). All of them provide data on education indicators and school participation in some form.

However, more than a decade since India launched its own programme of universalisation (SSA), and more than five years since it made elementary education a fundamental right (Right to Education, RTE 2009), it is still uncomfortably behind both international and national goals in education. Worse still, in the midst of this “feast” of data sources, we are not even sure how far we have reached with the goals of “Education for All” as we get varied, often contradictory evidence on basic indicators such as the proportion of children out of school, the extent of improvement in retention levels, learning outcomes and quality of education. Even in areas of education finance, such as teacher appointments and salaries, we do not have an authentic database. Further, the quality of the data collected is highly suspect, as evident from the fact that different sources provide vastly different estimates and the processes of verification and validation, are not in use. It would appear then that despite the fact that the coverage and scope of data collection has increased enormously with many more indicators added, and technology has enabled better management of data, some nagging questions remain about the quality, utility and purpose of the data, with obvious implications for planning and policymaking.

In addressing these issues, one is confronted with three sets of facts: (i) that the methods of data collection (including definitions of important indicators), of estimation and of data dissemination, appear to be inadequate and inappropriate to the purpose they are meant to serve; (ii) that the purpose itself, for which data is collected, and the use to which it is put (or should be put) are at odds with the larger goals of education; and (iii) the data are inconsistent across sources—both government and non-government sources—and even within government sources, raising questions not just about validation but also about the choice of data for different purposes.

This article is thus an attempt to unpack the methodological constraints as well as take a closer look at the administrative

anomalies that exist in the process of collecting data, which raise questions about the purpose for which data is collected and the use to which it is put. The latter in particular has a direct impact on the policy outcomes in the education sector. The article also looks at some of the major non-government sources of data to see if these have filled the gaps in government data and contributed to the goals of planning and policymaking in a positive way.

Methodological Discrepancies

With the advent of the SSA and DISE or U-DISE, a push was given to evidence-based policy and decentralised data and information. However, its import seems to have been lost on the government data system, which continues to be marred by loopholes in the processes for collection and dissemination of data with consequences on reliability and availability of data in desired forms and at desired levels of use. The discrepancies start with variations in the definitions and methods of estimation but also include aggregation errors, lack of verification and validation of data and other inconsistencies that end up yielding inaccurate information and vastly differing numbers across sources—government and non-government. With little effort made to clarify or reconcile the differences, the choice of data set influences the type of policy adopted. In fact, the use of non-government sources by policymakers has been on the rise in recent years, raising questions about the choice of data source by the government itself, and the import of this trend on the credibility of the government’s own system of generating information.

The methodological difficulties begin with the range of definitions and methods of estimation used for important indicators by different government and non-government agencies collecting data. For instance, estimates for out-of-school-children (OOSC),³ all collected through household surveys, are based on different “questions” being asked by investigators employed by each source. The NSS, for example, ask, “how many children are currently attending school?” While the census enumerators ask questions related to “status of attendance in an educational

institution.” The MHRD survey, on the other hand, claims to follow both the sampling and methodology used by the NSS, and yet arrives at vastly different results! The NSS and MHRD surveys that are based on a sample, then extrapolate from their figures, the proportion of children that are out of school, as a percentage of the population of children in that age group.

Using this method, the NSS 71st round has pegged the figure at a little less than 10% of the child population, amounting to nearly 20 million children, while the MHRD (SRI-IMRB) estimates put it at 3% and thus approximately three million! The Census 2011, on the other hand, suggests that more than 15% children in the same age group do not go to school, thus giving us a figure of almost 40 million!

The problems are compounded by the fact that formats are designed centrally and do not take into account local specificities and nor do they adequately train teachers—the primary data enumerators—to fill the formats. As a result even simple information like functional toilets may not be precisely enumerated. For instance, DISE, which is the only database collecting this information, provides no definition or even parameters for what “functional” denotes. It is left entirely to the understanding of the enumerators (schoolteachers) to make what they will of toilet functionality. To take another example, the figure for total number of teachers in a school turns out to be not as simple a statistic as it sounds, with teachers being routinely sent on deputation to other schools. Thus, it is unclear whether a teacher who is on deputation from another school, is to be counted in her current position or in her original school or does she end up being counted in both? Similarly, information on the employment status of teachers is more complicated and state-specific than a centrally designed DISE format allows. Most states hire teachers in different categories in addition to the permanent, such as voluntary, para, temporary, assistant and so on, not all of which are classified as “contract.” But the DISE format has only two categories—regular and contract, resulting in highly inaccurate

data being collected on an important indicator. Since each of the categories mentioned above is associated with a different salary structure, this information has huge implications for the education budget.

Data differences become even more problematic when it comes to measures of more nuanced indicators, such as those that encompass quality. This is never more glaring than in measures of learning outcomes. The two main sources of learning levels data are the ASER surveys (a non-government source) and the National Council of Education Research and Training's (NCERT) National Assessment Survey (NAS—a government source). However, the methodologies used by each are so different that comparisons would be meaningless. While ASER conducts tests of basic competencies in reading and mathematics at the household level, NAS conducts broader assessment tests at the school level. Much has been written about the validity and potential problems involved in testing at each of these places.⁴ In addition, ASER covers only rural schools, while NAS covers schools in rural and urban areas. But, ASER covers private unaided schools as well, while NAS covers only government- and private-aided

schools. Further, ASER collects annual data, but NAS does not. In other words the only two sources of learning data are so diverse in methodology that they cannot be compared or validated one with the other. Eventually, therefore, which data one chooses to use, assumes dimensions of personal or even political choice.

Another aspect of data credibility that has proved to be a weak link in the data collection process is verification and validation of data. While the rules for DISE dictate that 10% of the sample be randomly cross-checked, DISE itself is unable to verify that this process is either regularly or adequately carried out. This is due to lack of capacities available at the frontline for monitoring the process. Similar rules might well apply for other data sources, but the data collection agencies being highly centralised are unable to ensure that such verification in fact takes place. Hence, it would not be unfair to assume that data reliability is highly compromised by the fact that it is not cross-checked through a public verification process. The problem is made worse by the fact that the education departments use data collected under their aegis only, and ignore the evidence presented by other government

or non-government sources. In other words no process of validation is followed to improve the credibility of data collected. However, data validation is presented with mundane difficulties as well, related to different methods and time periods used for estimating different indicators. For instance, the Right of Children to Free and Compulsory Education (RTE Act talks about children between six and 14 years, but practically all data agencies (except those under MHRD) use different age groups when compiling education data, making it difficult to compare and validate. Table 1 shows the varied ranges covered by data sources on education alone.

Similarly, the dates and periodicity of data collection also vary across sources. ASER is an annual survey; the NFHS followed a six yearly pattern initially but has now slipped to 10 years since the last survey. The IHDS thus far has maintained a gap of six years between its two successive surveys. While the NFHS-3 and IHDS-1 are roughly from the same period (2004–05 and 2005–06), neither corresponds to the census dates, nor IHDS-2 (2011–12) does. The NSS also follows a different time period for its education surveys. Given that data collection is an expensive process, in terms of both time and money, if different agencies used comparable definitions and methodologies to collect data, we could have reliable time series data on education at limited cost. Sadly, that is not the case.

Table 1: Variation in Age Categories Used for Education Indicators by Various Sources

NFHS	IHDS	DISE	NSS	Census	MHRD (SRI-IMRB)
6–10 years	8–11 years	6–14 years	6–11 years	5–14 years	6–14 years
11–17 years	7–9; 10–14; 15–19 years		12–14 years	15–19 years	
15–49 years	7+; 12+ 17+ 19+ years		5–19 years	20–24 years	

NFHS: National Family Health Survey; IHDS: India Human Development Survey; DISE: District Information System for Education; NSS: National Sample Survey; MHRD: Ministry of Human Resource Development, Government of India.

Table 2: Government Data on Education

Survey	Organisation	Regularity	Variables
School surveys			
Selected education statistics	MHRD and state governments	Annual	Enrolment and teachers by grades, school type and management. State level data.
All India education survey	MHRD	Gap of six to seven years	Enrolment, teachers and infrastructure by grades, school type and management. State and district level data.
DISE and later UDISE data	MHRD and state/district governments	Annual	Enrolment, teachers and quality by grades, school type and management. State, district and school level data.
Household surveys			
Population census—enumeration	Central Statistics Office, Ministry of Statistics and Programme Implementation (MoSPI), Government of India	Decennial	Children attending education, institutions, education level attained and child work.
National Sample Survey Office (NSSO) education rounds—sample survey	MoSPI	Once in each decade	Children attending educational institutions, grades enrolled, education level attained, reasons for not attending, and costs of schooling.
SSA household survey—enumeration	State governments	Annual	Children attending school and out of school. State, district and village level.
State specific child tracking survey—enumeration	State governments	Annual	Children attending school and out of school, computerised and tracked over years. State, district and village level.
Out of school survey—sample survey	EdCIL (Education Consultants of India, Limited) and SRI—IMRB—commissioned by MHRD	Gap of three to five years	Children attending education institutions, grades enrolled, education level attained and reasons for not attending.

A primary reason why the available data sets have such a limited impact on policy is that the different agencies have planned their data collection for different (and specific) purposes, and not necessarily for planning or monitoring education and hence for education policy (Table 2, p 43). For example, the education rounds of the NSS are part of the survey on social consumption, which in turn is for the purpose of making an assessment of the benefits derived by various sections of society from public expenditure incurred by the government.⁵ The population census on the other hand is the primary source of basic national population data required for administrative purposes and for different aspects of economic and social research and planning.⁶ While the education ministry is not precluded from using any of these data for their own planning purposes, in fact they rarely refer to them. The non-government sources also have unique purposes in mind and not all of them have education as the primary objective either. Thus the NFHS is essentially a health and nutrition survey that also collects data on select education parameters. Similarly, the IHDS is geared towards the larger goals of human development and poverty, especially the links between education, skills and livelihood. Only ASER is solely dedicated to education, specifically learning levels. However, it does not tell us how the levels of learning vary with student enrolment or attendance, or any other household factor.

What is more surprising is that even the data collected by the MHRD and state education departments, though admittedly for the purpose of monitoring and planning, is not geared towards policy goals either. Instead, data collection and analysis are guided by its use in taking stock of the provisioning of schools, rather than as a mirror of their functioning. School surveys therefore focus unsurprisingly on collecting information related to (i) broad indicators of infrastructure and teacher availability; and (ii) student enrolment and distribution of incentives. Both these sets of data showcase administrative efforts rather than educational progress. Even the household survey (MHRD's SRI-IMRB) is

used only for estimating the oosc. No effort is made to use disaggregated data to understand problems of specific groups of children or schools.

Flawed Planning Process

A second conundrum with the purpose and use of education data relates to the fact that planning and policymaking are extremely centralised processes, based on the political and fiscal situation prevailing at the time, rather than the actual reality in the education sector. Thus, data however collected plays a limited role in the planning and policy processes. For instance, the Project Approval Board (PAB) at the MHRD that approves Annual Work Plans and Budgets (AWPBs) for the states does so on the basis of the finances allocated to it from the Ministry of Finance (MoF) and the norms of expenditure specified by the central ministry, that is, the MHRD. While the AWPBs reflect the needs of each state, eventual allocations differ widely from it, as they are based on what is made available by the MoF through processes that do not involve the education sector. Of course, state plans are themselves based on a process of aggregation that does not involve a genuine decentralised planning process. This is evident from the fact that dissemination strategies are not aligned with the goals of decentralised planning, as data is largely unavailable in usable form at the local or school level. In fact, local data management systems are virtually non-existent, putting paid to the idea of decentralised planning. Thus, while it is true that schools are now asked to prepare their plans through the School Management Committees (SMCs), in fact, what is submitted by them are copies of the DISE format—presumably as indicative of the status of schools and thus reflective of their needs! Eventually, therefore, at the district and presumably at the state level as well, DISE data is referred to for making the state AWPB. Given the discrepancies and problems cited above with DISE, this is clearly a flawed process. What is perhaps more unpalatable, however, is that once allocations are made at the PAB, their distribution down the line to each school is a completely opaque process, likely guided by the level of political influence

enjoyed by a district or even a sarpanch who is the head of the panchayat. Here again, facts, data and so on, appear to play a negligible, if any, role.

Therefore, the national-level data sets, such as DISE, serve a descriptive purpose more than a functional one—at best being used for making comparisons between states. Little use is made of the data for understanding why progress is different or less than expected or what the location-specific problems are or how disaggregated data can be used to address them. This process in fact makes the objective of a decentralised data system feeding into a process of centralised planning, oxymoronic.

Limited State Capacity

The third and perhaps overarching problem that faces the data regime in education (and possibly other sectors as well) is that of extremely poor and limited capacities to design, collect, analyse and use data throughout the government structures from the central to the local. DISE is run almost entirely on the shoulders of data entry operators of the education departments at the district and block levels. At the central level it is housed within the National University of Educational Planning and Administration (NUEPA) with a bare staff of two—none of who are statisticians—and a set of technical personnel who operate the sophisticated digital structure for aggregating and then disaggregating data sent up from the states. Further, data that is collected from the ground up amounts to a process of simple aggregation resulting in the loss of specifics, such that by the time it reaches the central level, it barely reflects the ground realities and could hardly serve the needs of the people on the ground. The aggregation itself is still done manually at the block level in many states with digitalisation appearing only at the district level. Then too, checks and balances are few with the bulk of the responsibility falling on the limited number of data entry operators available in each state.

In addition, the departmental staff at the state level have not acquired the capacity, through their own qualifications or through government provided training

on relevance and importance of quality data or its use in the planning or policy process. For instance, innumerable formats are designed for monitoring schools, but none of that data is put to any use.⁷ In fact it is not even referred to in the monitoring review meetings held at the block and the district. Unfortunately, the personnel involved in collecting and collating that information are themselves unable to gauge the importance of that information and so are unable to use it in a constructive fashion. One of the ways the data from monitoring formats could be used is to cross-check with DISE data (or vice versa) to improve the quality of both data sets. But data collection is seen as a simple job of “filling formats”—its import is completely lost on all personnel in the department. Hence the time and effort devoted to a process of generating information is virtually lost.

Instead, implicit in the collection process is a conflict of interest, especially with DISE data—now the official database for education—as it is entirely dependent on formats filled by teachers. It is well established that teachers might be incentivised to represent information in ways that inflate facts, such as student enrolment, for at least the following reasons: (i) to indicate that they have succeeded in the enrolment drives they are expected to carry out at the start of each school year and on which, their own employment might depend; (ii) to elicit greater funds for mid-day meals, which are allocated on the basis of enrolment. Teachers are also often under pressure from parents to add or keep the names of their child on the rolls even when they are not attending school or are absent for long periods.⁸

It would appear then that the data regime is constrained by the fact that neither is it fully versed with the ground realities, nor fully served with resources to design and supervise the process. It is also not in conversation with each other to harmonise the imbalances and improve the statistical validity of the data. Thus, despite their best efforts and intentions, the quality of data generated is not entirely reliable. The process is further hampered by the fact that the

purpose for which data is generated remains at odds with both the goals of planning and policy as well as eventually the goals of improving the quality of education itself.

Non-government Data

It must be pointed out at the outset that contrary to what might be expected the non-government data sets do not take off from where the government data sets end, at least not in terms of addressing the gaps in government data (Table 3). In fact the trend of non-government agencies collecting large data began in the 1990s, when the government data also began to diversify and expand. While the addition of some new indicators did provide an opportunity to extend the use of data beyond the government sources, as mentioned above, that has not happened except in a few select cases.

First, while new indicators have been added by the non-government agencies, the indicators used by each of them are dissimilar to each other. For instance, the NFHS collects enrolment, attendance and attainment data, while the IHDS collects literacy, dropout, learning levels, skills and expenditure data. While ASER collects learning levels data, like the IHDS, it uses a different method. In fact, the methods used are different from the government’s methods in the National Achievement Surveys as well.

Second, the periodicity and age categories considered by non-government data sets are also as variable as the government data sets. Hence, validation or

comparability is not possible within these data sources either, adding to the confusion, rather than providing greater clarity.

Third, not all of them are geared towards policy advocacy either. For instance, the IHDS data is largely for academic purposes and disregarded by the education department as far as policy is concerned, as is the NFHS data. ASER on the other hand has been extensively used to influence policy. Interestingly though, ASER data does not include information on school participation, neglecting a crucial aspect of the context within which the indicators are placed.

In addition to ASER, select research studies, which have emerged from academic quarters have also had a greater influence on policy than the large data sets generated by the government. Even though these are based on much smaller samples and look at a limited set of variables, their impact on policy has been disproportionately high. One possible reason for this could be their public dissemination strategies, which have been more forceful than those of other agencies leading to the wide influence they have had. Another is the disproportionate value attached to academic work, especially that based on quantitative studies. This is perhaps best evidenced by the influence of the randomised control trials (RCT) methodology and the studies it has spawned, which have been increasingly at the forefront of policy advocacy in education in the last decade or so. Their impact must be seen against the fact that they filled a critical gap

Table 3: Non-government Data on Education

Survey and Organisation Involved	Age Categories	Regularity and Year of Survey	Variables
Household surveys			
NFHS-I, II, III	15–49 years (for women) and 15–54 years (for men)	Gap of 6 years (1992–93; 1998–99; 2004–05)	Enrolment, attainment levels, attendance
NFHS-IV		Gap of 10 years (2015–16)	
International Institute of Population Sciences			
IHDS-I and II	5+ years	Gap of five years (2005–06 and 2011–12)	National and state literacy, enrolment rates; never-enrolled, dropout rates, repetition/failure; absence (six+ days); reading, writing and arithmetic levels; skill levels for 15–49 year olds
NCAER and University of Maryland, US	7+ years 5–14 years		
ASER	3–16 years	Annual (1999–2014)	Learning outcomes in literacy and numeracy
ASER centre and PRATHAM			

Source: NFHS-3 (2005–06); IHDS-1 (2004–05); NSS (71st round, 2014); SRI–IMRB (2014).

in the data and analysis system within education. However, they have acquired the status of a “gold standard” amongst researchers and policymakers to the exclusion of all other methods. But quantitative data sets that do not provide information on processes and other qualitative indicators can hide more than they reveal, presenting an inaccurate or incomplete picture that could bias policy. Unfortunately, qualitative studies that focus on contexts and address questions starting with “why,” struggle to find space and voice in policy circles. The choice of data thus acquires a larger implication and raises important questions of subjectivity and data neutrality in the process of collection as well as interpretation and use of data.

Conclusions

What emerges from the above discussion is that while an improved data system cannot by itself guarantee a greater impact on planning, monitoring and policymaking as that requires building capacities for local data management, mechanisms for dissemination and public verification, as well as changes in the processes of planning and policymaking—which are harder to execute—sorting out the numbers game is a necessary step towards it. In other words, the criticality of good quality data for governance must be appreciated in the state machinery and steps taken towards ensuring its availability.

While the quality of evidence and data is of concern in itself, the problem is made worse by the differences in public dissemination strategies of data by sources and their impact on policy. While non-governmental data sources have played a particular crucial role in recent years, it is worth stressing that their impact has been not just disproportionate but also the subject of some controversy in recent years. Individual research studies can shift the terms of the public and policy discourse by highlighting critical facts in an effective and robust manner, but their use in public discussions, policy notes, and by other researchers has given them the status of accepted wisdom and drivers of policy pronouncements to the detriment of

other sources of wisdom and other voices from the field. This trend must also be guarded against.

In this context, the article sounds a cautionary note against subjectivity in the use of data. Data is not neutral as it is collected through interaction of the data collector and the respondent through varied tools of data collection—so bias is likely to be introduced from all three elements. As a way forward, therefore, the article makes a few suggestions: (i) a greater role be envisaged for data users, especially the education officials at different levels of government ranging from the national to the local. Necessary technical skills, if provided to them, can enable them to be cautious when collecting data as well as enable them to interpret and use it appropriately such as when making their plans; (ii) similar support could also be provided to other monitoring agencies, social audit groups, the school management committees and education researchers to allow them to publicly verify data that is officially collected; and (iii) bias can also be estimated through the use of multiple data sets. Data, especially when used for policy if validated against different sources, can ensure that it factors in bias and is therefore used more judiciously.

In fact, multiple data sets have other uses as well. For instance, we know that any single data set cannot collect information on all relevant issues, because data collection is a very expensive and time-consuming process. Thus information collected by the NSS on household expenditures, which demonstrates that 70% of all the OOSC in urban areas are concentrated in the lowest quintile, while in rural areas they are in the lowest two quintiles, is relevant information that can and should be used by the education department without having to repeat the exercise. Similarly, the NFHS data provides linkages between education participation and family health, also of importance to the education department.

Making better use of data requires proactive collaboration of different government and non-government agencies. For instance, if household and school data were available in the same portal, it would maximise their use. Similarly, if

the NSS education rounds were better coordinated, along with standardisation of definitions of important indicators, it would greatly help in serving the cause of education goals. In this regard streamlining the planning process to enable planning based on decentralised data will go a long way in improving the use of data at the local level as well as ensuring a more genuine decentralised planning process.

These issues have assumed greater importance today as we stand on the cusp of deliberations on a New Education Policy. With the range of data sources available, one wonders which will be used by the education administration. As stated by T C A Anant, Chief Statistician of India, in a meeting on education data, “it is important to work out the uses of data for policy and the corresponding indicators, before a policy is framed, rather than after the policy is put into place.” This simple advice has unfortunately not been followed thus far, resulting in some of the problems seen in the use of education data for policy purposes.

NOTES

- 1 The report, which was released in 1999, covered Uttar Pradesh (undivided), Rajasthan, Bihar, Madhya Pradesh (undivided), and Himachal Pradesh. The last state was an “outlier” in the set as it revealed a wholly different story of state action in education.
- 2 U-DISE or Unified-DISE is a database of all students from grades 1–12.
- 3 Non-government sources do not collect information on this variable at the national level.
- 4 Krishna Kumar (2015), “We Need a Real Learning Grid for India’s Elementary Schools,” *Hindustan Times*, 21 January.
- 5 <http://mail.mospi.gov.in/>.
- 6 http://censusindia.gov.in/Data_Products/Library/Indian_perceptive_link/Census_Objectives_link/censusobjectives.htm.
- 7 See Bhatti, K and R Saraf (2016), “Does Government’s Monitoring of Schools Work?: A Study of Frontline Education Bureaucracy in India,” CPR Working Paper.
- 8 See Bhatti, K, R Saraf and V Gupta (2017), “Out of School Children—Some Insights into What We Know and What We Do Not Know,” *Economic & Political Weekly*, 9 December, Vol 52, Issue No 49.

Economic & Political WEEKLY

available at

People Book Stall

15, Cawasji Patel Road

Fort, Mumbai 400 001

Ph: 22873768