

# REVITALISATION OF EDUCATIONAL STATISTICS IN INDIA

## Issues and Strategies

**Yash Aggarwal**

Senior Fellow and Head, ORSM Unit, NIEPA, New Delhi

### Contents

1. Need for educational and allied data
2. Data collection formats and frequency of data collection
  - 2.1 *Educational statistics: A historical perspective*
  - 2.2 *The Statistical Commission*
3. Design of information systems for educational management
4. Database on elementary education in India
  - 4.1 *Types of educational data*
  - 4.2 *Type of allied data*
5. Sources of educational and allied data: National/state level
  - 5.1 *Frequency of data collection and reporting*
  - 5.2 *Data flows and aggregation*
  - 5.3 *National data system at MHRD*
  - 5.4 *All India Educational Survey*
  - 5.5 *Census of India*
  - 5.6 *National Sample Survey (NSS)*
  - 5.7 *Planning Commission*
  - 5.8 *Survey of All India Educational Administration*
  - 5.9 *State Level Educational Statistics*
  - 5.10 *Other Sources*
6. Sources of educational and allied data at the district level
  - 6.1 *District education office*
  - 6.2 *DISE database*
  - 6.3 *School census*
  - 6.4 *Social and economic data*
  - 6.5 *District census handbook*
  - 6.6 *NICNET*
  - 6.7 *Total literacy campaigns*
  - 6.8 *Other sources of data*
7. Limitations and inadequacies of the data collected
  - 7.1 *Conceptual and definitional problems*
  - 7.2 *Useful data wasted*
  - 7.3 *Lack of coordination among data collection agencies*
  - 7.4 *Data gaps*
  - 7.5 *Time lag*
  - 7.6 *Reliability and redundancy of data*

*7.7 Developing comparable time series data*

*7.8 Linking data across populations and levels*

*7.9 Weak capacity to analyse and interpret educational statistics*

8. Recent attempts to revitalize educational statistics

9. DISE database and its features

10. Future possibilities and recommendations

11. Summing up

---

## 1. Need for educational and allied data

Human beings are at the centre stage of all development activities. Most of the statistical systems are designed to capture the interface between the development programs and various aspects of the human well being and their quality of life. Education is one of the largest social sector development programs. Seen in this context, there is a growing body of literature that suggests that the basic necessity for the success of any sustainable reform is to have a continuous and unbiased assessment of the system by clearly identifying its strengths and weaknesses. No development program has ever been completed successfully without a sound monitoring and decision support system. The slow progress or the failure of a program to take-off is often associated with poor monitoring and feedback mechanisms.

To co-ordinate various management functions and to improve planning processes, not only a variety of data is needed but is needed in a form and at a time that is consistent with the requirements of user agencies operating at various geographical and administrative hierarchies. The complexity and diversity of decision support system increase directly in proportion to the outreach and physical deployment of movable and immovable resources in geographical space. Education is a unique development sector with the highest degree of penetration in geographical space as compared to any other organised activity under the management and control of the government. It employs largest number of educated manpower spread all across the country. These characteristics of the education sector make it necessary to have strong inter-linkages among the data elements, flow of information and their sharing across various users. As India strives to increase access and participation of individuals in various types of educational activities as a part of the life long learning activities, the complexity, diversity of courses offered and spatial spread will increase manifold. Similarly, the demand for information multiplies as the levels of decision-making, control and monitoring increase. The move towards decentralisation of decision making in social sectors, including school education, will not reduce the need for centralised decision support and monitoring systems but will necessitate the development of integrated and national level information systems so that the information/data needs of development administrators at various levels could be met in an integrated, consistent, comprehensive manner.

Management and monitoring of the progress of social development activities require data on key variables not only relating to a particular sector and its sub-sectors but from a variety of sources allied to the sub-sector. For example, management of primary education requires data on population, settlement structure, social stratification, social

and economic development variables having direct and indirect bearing on strategies to improve access. In addition to primary education statistics on access, retention and quality of learning, additional statistics such as those on child health, mortality, immunisation, maternal mortality and malnutrition are critical to measure the well being of the children. The latter set of indicators also influences the outcome of primary education reforms. Similarly, the employment status of the family has an everlasting influence on the health and educational status of the young children. All these variables are inter-related in a complex manner. The effort here is not to develop a model of educational planning but to highlight the inter-linkages within the social sectors.

The pay-off from an improved, consistent and timely availability of data on key and simple performance indicators of social and economic development can be enormous. It will not only provide the right type of signals to development planners, but also help in more sharply focusing the scarce resources for areas/activities where these are most needed. Educational planning is no exception. Therefore, like other sectors of the economy, a modern and well staffed system of data collection, collation, analysis, communication, sharing and dissemination for the education sector as a whole and for its various sub-sectors is a prerequisite for planning to be effective and responsive to the needs of the stakeholders.

Broadly speaking, the education and allied data is needed for the following purposes:

- To understand the historical processes of evolution of educational, social and economic growth and to analyse the disparities in international, intra and inter-regional, educational, social and economic development.
- To understand the nature of spatio-temporal relationship, association, similarities and dis-similarities among the selected educational, social, economic and related indicators of human development.
- To build models, undertake simulation and projection exercises based on past trends and/or to evolve feasible solutions based on specified future targets/ objectives.
- To monitor and evaluate the fund flow and their utilisation pattern, trends in quantitative and qualitative characteristics of certain activities/schemes/events through selected indicators so as to provide feedback on the degree of success and to evolve suitable intervention strategies consistent with the stated objectives and to undertake midcourse corrections, if necessary.
- To provide decision support services to educational administrators, school Principals/ head teachers, community leaders, students and teachers to exercise the rational choices and promote the process of norm based decision making based on the most recent data and to permit comparative analysis of performance of their school/area with other schools/areas and groups of population.

- To share data among other users and interest groups by developing an institutional network for dissemination of relevant information.

Historically, India is one of the few countries where a fairly advanced machinery of data collection for various social and economic sectors existed much before the independence. The population enumeration started as far back as 1872 and has been a decadal exercise since then. India is known for completeness and exhaustiveness documentation of social, economic and political developments. The collection of key economic and social statistics started during the middle of nineteenth century and by now a fairly large number of time series data on social and economic indicators is available. The successive governments continued to strengthen the historically obtained system of data collection and documentation for economic and social sectors of the Indian economy. The existence of district level Gazetteers is the finest example of the efforts to compile the social and economic data at a disaggregated level. The revenue records of India present another example of a historically evolved system of data collection and its use by district and state administration for revenue mobilisation. Education was no exception. The educational statistics which were collected till fifties and early sixties were of superior quality and much more elaborate than what is available today. In the last few decades, there has been a progressive decline in the scope, coverage and reliability of educational statistics and data collection on some important aspects was even discontinued without evolving alternative mechanisms for its estimation.

## 2. Data collection format and frequency of data collection

India has a long tradition of developing a framework for recording of educational statistics. These statistics were used for funds flow management, teachers recruitment and deployment and administration of government's control over the educational system. As the following discussion would show that the data collection could not keep pace with the increasing demand for data.

### 2.1 Educational statistics: A historical perspective

India is perhaps one of the few countries which established a sound statistical system centuries ago. In the case of education statistics, '*Educational Development in India*' dates back to 1870 (MHRD, 1999). This was followed by the introduction of a more detailed statistics by Indian Education Commission on quinquennial basis for 1886, 1932-37. A decennial view was presented for the decade 1937-47. Annual educational statistics began to be collected from 1913-14 followed by elaborate quinquennial reviews. Prior to 1947, the educational statistics was collected by the Directorate of Commercial Intelligence and Statistics in form S-163. This activity was taken over the Ministry of Education after the attainment of independence in 1947. This form was replaced by form 'A' w.e.f 1949-50 containing comprehensive and detailed educational statistics issued in the form of *Education in India Vol. 1 and Vol. 2*. Due to time lags in the collection and compilation of information contained in form 'A', it was replaced by form A, A1, A2 and A3. Form A contained the minimum essential data, A1 contained data in respect of SC/ST; A2 contained supplementary data for planning purposes and A3 contained districtwise data on quinquennial basis to be collected from 1964-65

onwards. However, the simplification and introduction of more forms did not help and the time lags continued to increase. Consequently another revision of forms took place in 1976-77 and a new set of forms under ES series were introduced.

ES-I	Numerical data
ES-II	Financial data
ES-III	Examination results
ES-IV	Data in respect of SC/ST
ES-V	Districtwise data
ES-VI	Information on special studies

Introduction of new formats is not enough to resolve the problems of the data collection, management, analysis and their use. The problems relating to the availability of data continued to persist and the efforts of the central government did not yield any positive result. Subsequently, the ES series of forms were further bifurcated for the schools and higher education separately. Yet another revised version of formats were introduced from 1984-85 onwards.

ES-I(S)	Numerical data (school education)
ES-II (S)	Financial data school education.
ES-III	Examination results-school and higher education.
ES-IV (S)	School educational statistics for SC/ST
ES-I (C)	Numerical data in respect of higher education
ES-II (c)	Financial data for higher education
ES-IV (c)	Numerical data for higher education of SC/ST

Despite all the reforms in streamlining the collection, collation, analysis and availability of educational statistics, the situation worsened from bad to worse. While most of the vital information was not available, even the availability of routine information became a victim. The problems with the availability of statistics now are worse than what it was twenty or thirty years ago. This clearly reflects the gravity of the situation and limitation of the strategies to overcome the problems of educational statistics in India. It also shows that the traditional approaches have failed to deliver the results. Therefore, an overhaul of the system will be required as far as the collection and availability of educational statistics in the country is concerned.

India has recently joined the UNESCO/OECD Pilot project on World Educational Indicators (WEI). The OECD has identified 43 indicators for collection of data and covers the following areas:

- a) Demographic social and economic context of education
- b) Cost of education on human and other resources;
- c) Access to education, participation and progress in education;
- d) School, environment/school and classroom processes;
- e) Graduate output of educational institutions;

The detailed analysis of the indicators corresponding to the above categories would reveal that India at the present stage is ill-equipped to generate meaningful data on most of the WEI. The more worrying is that fact that after having agreed to participate

in the project, no efforts were made to have a realistic assessment of the states' ability to collect and report on the data on the new indicators. Can the existing system respond to the emerging national and international needs? As of now, the answer is definite no.

Merely asking the states to provide data without any serious efforts is expecting too much with too little effort. It is rightly characterised as 'low priority and high expectation exercise'. The experiment is bound to fail as many earlier efforts have failed and then the bureaucrats will blame it on the system and its preparedness. Why an objective assessment of the situation can't be made before taking a plunge. The following factors were responsible for deteriorating quality of educational statistics in India and their resolution is necessary to bring the desired improvements in the relevance and quality of educational statistics.

- Even till today, there is no vision and long term perspective on the role and objectives of educational statistical system.
- Inability of the educational planners to emphasise on the role of norm based planning and practices in budgeting, planning, monitoring and evaluation.
- Low priority accorded to modernisation of data collection, management and analysis.
- Inability of the educational administrators to come out with agreed set of indicators for monitoring and measuring progress towards the achievement of goals. Even after decades of development planning, it is not how a district or a state is to be ranked in terms of educational development.
- Lack of standardization of definitions and concepts used during data collection and statistical analysis phase. The diversity in educational systems further complicates the analysis and interpretation of educational statistics.
- There are no clear guidelines/manuals on educational statistics in India which could be used by data collection agencies. Similar guidelines for data processing and analysis are missing.
- Lack of a forum to provide continuous and on the site support for data collection and analysis to the field staff and also for the senior educational administrators engaged in planning and management of education.
- Indifferent attitude of the states to participate in the national data collection system. The central government relies on the state governments for data collection and reporting. Many states governments see this as a routine activity and involve the staff in other types of work neglecting the educational statistics work. The centre has not been able to check this malaise. Statutory provisions regarding collection and supply of basic educational data are required. Panel provisions for supply of correct, valid and timely data do not exist in India. Of late, the Central Advisory Board of Education has not been fully functional. The recent stand-off between the central government and many states government on matter of text-books has further complicated this issue.

- The states do not accept the responsibility for the quality and reliability of educational statistics. It even becomes difficult to ascertain the quality index of educational data. There is no system of quality assurance or post-enumeration checks. This becomes important since all the educational data is collected through teachers. It is important to validate the quality and coverage of administrative data.
- Complete secrecy about educational statistics. The data at school/district level is not shared with the stakeholders. The backward flows of data are totally absent. All data remains locked in the offices of the DEOs and the state Directorate of Education.
- Redeployment of statistical staff, at the district and the state level, for other functions has hampered the smooth collection and flow of data from one level to another.
- The bifurcation and trifurcation of directorates created its own problems of co-ordination and collection of educational statistics from different type of institutions.
- Absence of electronically managed data bases for monitoring purposes has caused considerable time and cost over runs even in the projects and schemes of national importance.

Practically all development sectors are faced with the situation similar to the above. The efficacy of many social welfare schemes can not be assessed for want of crucial data. In fact, educational statistics is one of those sectors where adequate attention has not been paid to revitalise the data collection, analysis and dissemination machinery at various levels. Data from two or more source seldom match and differ to unacceptable levels.

The central and state governments taken together spend about 13 percent of their budget on education and allied activities with no system of monitoring. A sum of Rs. 327 billion was earmarked for education in the budget of 1994-95. It is difficult to estimate as to what proportion of this budgeted amount was spent on monitoring and strengthening of educational statistics. In all probability, it will be a negligible amount intended to maintain status-quo. A radical departure will be necessary to re-establish the system of educational statistics through a complete transformation of the processes of data collection, collation, flows and analysis of educational data, on the one hand, and by necessary investment in modernisation and capacity building, on the other.

## ***2.2 The Statistical Commission***

The Government of India appointed a Statistical Commission in year 2000 to examine various issues related to organizational structure of the statistical machinery in various departments, status review of the existing data base on the Indian economy and also to suggest changes modification to meet the emerging challenges for improving the efficiency and effectiveness of statistics for decentralized planning and management. The Terms of Reference for the National Statistical Commission were as follows:

- a) To examine critically the deficiencies of the present statistical system in terms of timeliness, reliability and adequacy;

- b) To recommend measures to correct the deficiencies and revamp the statistical system to generate timely and reliable statistics for the purpose of policy and planning in Government at different levels of administrative structure;
- c) To recommend permanent and effective coordinating mechanism for ensuring integrated development of the decentralised statistical system in the country;
- d) To review the existing legislation for the collection of statistical information and to recommend amendments, where necessary, to achieve the objective of collection and dissemination of timely, reliable and adequate statistics;
- e) To review the existing organisation of the Ministry of Statistics and Programme Implementation (Statistics Wing) and other statistical units of the Government and to make recommendations on their staffing and training requirements to enable them to cope with the increase and development of statistical services;
- f) To examine the need for instituting statistical audit of the range of services provided by the Government and local bodies and make suitable recommendations thereof; and
- g) To recommend any other measures for improving the statistical system in the country.

The final report of the Commission was submitted in July, 2001. While reviewing the status of social sector statistics, the commission noted that 'there are major deficiencies of data, which can be largely attributed to the near collapse of the Administrative Statistical System. The deficiencies common to all the sectors include: poor quality of data collected by the statistical system, inordinate delays, lack of effective checks, incomplete coverage, inconsistent data, poor implementation of provisions of Acts, low priority and general apathy to statistical activities, inadequate infrastructure & staff for statistical work, and lack of computerisation & its use in data compilation, processing and dissemination of data produced by different agencies. As a result, routine data on schools, students enrolled, hospitals, medical and para-medical personnel, births and deaths occurring in the population are just not collected due to a lack of proper emphasis on these items of information and the administrative back up for a compilation and analysis of the required data'.

- Unit-level data of one department would become accessible to other departments;
- A central storehouse of unit-level data of all departments should be created in the DES;
- Standardization key data elements should be attempted for use in all departments (such as permanent village codes) that should be included in all forms of data collection.

Though "Statistics" is under the Concurrent List of the Constitution and "Surveys" is only under the Union list, it is more important that the collection of statistics on any subject vests in the authority (Central Ministry or State Government Department) that is responsible for that subject according to its status in the Union, State or Concurrent Lists in the Constitution of India. When the proposed National Commission on Statistics (NCS) determines certain statistics as Core Statistics, in deciding on the agency that should be responsible for their collection, it will have to do so in consistence with the distribution of subjects in the three Lists. Therefore, when



proposing legal measures for Core Statistics, the NCS may have to propose different Acts for different subjects according to the List to which the subject belongs.

The statistical commission after an exhaustive review of the existing legal basis for collection of administrative and other statistics has come to the conclusion that the existing laws are highly inadequate to meet the diversified needs of educational and allied data. It has proposed a new legislation on statistical systems. Necessary legal provisions should be made, either by expanding the scope of the present Collection of Statistics Act (1953) or by passing new Act or Acts to:

- Cover any topic under Core Statistics, as defined by the proposed NCS;
- Make it obligatory on the part of individuals, or enterprises, or State and private agencies to provide the information sought for any survey under the aegis of the NCS;
- Provide right of access to records, including the record of Government agencies for statistical purposes;
- Ensure the informant's right to privacy by making it illegal to publish the identity of the informant, or by requiring him to furnish sensitive information;
- Provide penalties for informants, for their refusal to supply, or for wilfully supplying wrong information;
- Make it a penal offence for a statistical officer authorised to collect, process, or disseminate information collected from any survey under the Act, to wilfully distort or manipulate the data.

It is expected that during the Tenth Five Year Plan, the government would come out with a position paper on the recommendations of the commission.

### 3. Design of information systems for educational management

The first step in building a comprehensive indicator system is the development of a conceptual framework based on policy objectives, programs and the interests of policymakers and educators. Models of educational management should be such that the indicators selected are comprehensive enough to adequately describe the important dimensions of the education system, explore relationships among elements, and address issues of monitoring, evaluation and periodic review of policy perspectives. A comprehensive and consistent set of indicators must be flexible in terms of scope and coverage and based on a model of the educational system that is sufficiently general, yet complete, that its validity will continue over time.

There are different agencies managing different levels of programs and educational activities. While the directorate of education in each state are primarily responsible for the collection and analysis of educational statistics, their structure and line of command differs from state to state. For example, in West Bengal, there are as many as five ministers who deal with various levels of education. Similarly, separate directorates have been established for primary, upper primary and secondary education. There are also separate directorates dealing with non-formal and adult education programs. Generally, affiliated colleges offering general degree courses in Arts, Science and Commerce are controlled by the affiliating university and are funded

by the state government. The funds for the development of higher education are channeled through University Grants Commission and the state governments. Central universities are directly controlled by the University Grants Commission. The colleges of technical and professional education are also controlled by different universities. The Polytechnics offering diploma in various streams of engineering are managed and controlled by the Directorate of Technical Education. The Industrial Training Institutes are controlled by the Department of Labour and employment of the state governments. The Indian Council of Agricultural Research controls and funds the development of agricultural universities in India. Funds for the development of medical education (para-medical and degree level professionals courses) are provided by the Ministry of Health and Medical Council of India. It is therefore, necessary that a set of core variables should be defined for various levels of education so that a comparable set of data is available from the concerned Ministries/departments.

Realising the importance of information systems, considerable emphasis is given to the promotion and introduction of computerised educational management systems around the world. Computerisation is necessary because it offers tremendous economies of scale in storing, retrieval, analysis and sharing of data among spatially dispersed units of educational management. Countries of the size of India have no choice but to go in for large scale introduction of computers in educational management. The direct cost of establishing such systems will soon be off-set by the indirect costs due to poor management practices, inefficiencies in resource deployment and delays in non-availability of mission critical data. It is thus apparent from the above discussion that the data needs for the educational systems arise from the following considerations:

- Education policy objectives in general and those of the development perspective in particular.
- Organisational structure and reporting systems at various levels.
- Focus of various program interventions in terms of coverage and targeting.
- Level and degree of decentralisation in decision making.
- Differentials in spatial spread of infrastructure and manpower deployment.
- The management functions to be performed at each level.

The classical model of educational planning follows input-output approach. Consequently, the following grouping of indicators is considered appropriate for the design of information systems.

- Measuring access and participation: Measuring access to educational facilities and participation is the starting point of any educational system. Different state governments follow different norms for opening and upgradation of schools and colleges. Access is measured in terms of intake rate and availability of schooling facilities to all children in the relevant age group. It is important to distinguish between the availability and utilization of educational facilities. The availability is a necessary condition but may not be sufficient for its utilization by all social and economic groups.

Similarly the schools also receive various types of inputs from time to time. These include teachers, additional facilities and equipment, instructional materials, teaching aids, incentives for distribution among students etc.

The main variables/indicators of access, participation and input are:

- % children having access to primary schooling (formal and non-formal).
  - % habitations without schools as per the prescribed norms.
  - Availability of non-formal or other modes of education in unserved habitations.
  - Proportion of urban schools to urban population.
  - Intake rate by gender and social groups.
  - % students in private schools to enrolment of public schools.
  - Gross and Net Enrolment rate.
  - % classrooms in good condition to be measured through an index of the quality of building infrastructure.
  - % schools not inspected.
  - Classroom space per student.
  - School size by category of school.
  - Teachers and their distribution by gender and caste.
  - Schools without single female teacher.
  - School facilities like blackboards, chalks and dusters.
  - Availability of playgrounds, toilets, boundary wall and water etc;
  - Children receiving incentives of various types.
  - % education budget to total budget of the centre/state governments
  - % budget allocated for elementary education.
  - % GNP spent on education.
  - Ratio of private to public expenditure on education.
  - Number of administrative and inspection staff per 1000 schools.
  - Average number of in-service training days for a teacher.
- Process and efficiency related indicators: It refers to the interactions of various inputs to produce outputs. The resource utilization will also be addressed through process related indicators. Main indicators of process are:
    - Drop-outs and wastage by gender and region.
    - Repetition rate by gender, class and region.
    - Grade transition rate.
    - % schools with multi-grade teaching.
    - Utilisation of the educational infrastructure and provisions.

- Number of school working days in a year.
  - % expenditure on non-salary components.
  - Average attendance of the students.
  - Average days of teacher's absenteeism.
  - Pupil: teacher ratio.
  - Students per class.
  - % students reaching the final year of primary education cycle.
  - Teaching learning methodology.
- Output indicators: Outputs are the results or products of the education system. The number of successful graduates is one important measure of the outcome. A good measure of output is the completion rate (Aggarwal, 2002). In the recent years, emphasis is also placed on learning achievement. How many children have succeeded in achieving the desired level of competencies? The following indicators are included in this category:
    - % children who eventually complete the primary cycle in five years.
    - % children who attain the prescribed competencies beyond a specified threshold.
    - Input-output ratio.
    - Completion rates by gender and social categories.
    - Average years taken to complete primary cycle.
    - % Children who complete the primary cycle in the prescribed years of schooling and with prescribed levels of educational attainment.
    - Average cost per graduate.
    - % children getting admission to next level of education.
- Contextual and social and economic indicators: This category of indicators refers to all dimensions of the immediate and wider concern to the functioning of the school. Examples of context variables are the socio-economic status of the school population and general development scenario. These include:
    - Per capita income
    - % population below poverty line.
    - % population of socially and economically disadvantaged groups.
    - % population engaged in non-agricultural occupations.
    - Workforce participation rate for male and females.
    - Adult literacy (males and females)
    - Average number of years of educational attainment.
    - Index of gender disparities in educational attainment.
    - Index of social disparities in educational attainment.

- % Population in smaller habitations (population less than 300).
- % Population having access to safe drinking water.
- % Population having access to preventive health services.
- Radio and television sets per 1000 population.
- Newspapers circulation per thousand population.

The indicators discussed above normally are understood in the sense of cross-sectional database. It is also important that the same type of indicators would be available in the form of time series so that temporal changes and growth rates of various indicators could be examined. The temporal analysis of educational indicators is needed not only to assess the progress made but also to ascertain that the progress does not lead to the accentuation of gender, social and regional disparities. Therefore, disaggregated indicators across gender and social groups constitute an important dimension of information systems' design.

While the above list of indicators is general, the final choice of a set of measurable indicators should be governed by educational issues emerging from policy, implementation and feedback obtained from stakeholders. While the advanced countries have further refined the above indicators and added a few more, the developing countries find it difficult to even generate data of high quality and make it available quickly for selected few indicators belonging to each of the above categories.

At the primary and middle level of education, most public money is spent either directly by governments or transferred to educational institutions to acquire resources. The private aided recognised schools also receive matching grant from the government. Nevertheless, the data shows that the government share in expenditure on primary education has increased considerably since in the early fifties. The latest data suggest that the government is meeting about 86% of the total expenditure on education. While the corresponding data for primary education is not available, the share of government is much larger at the primary level as 97 recognized percent primary schools are funded by the central/state governments.

Considerable progress has been achieved in international comparability and reporting of educational statistics. The UNESCO evolved International Standard Classification of Education (ISCED) and is being followed for reporting of internationally comparable data for the member countries. Despite all the improvements in the recent years, the classification suffers from many limitations. Important gaps which make international comparisons hazardous include the classification of education by different levels, the definition and concepts of educational institutions, part time and full time classification of students, enrolment, teachers and other related indicators of internal efficiency. The diversity of education systems and differences in the structure of governance of education provide a challenge to international comparisons. A consistent classification of types of educational programs is yet not available. The current distinction between vocational and general education poses problems of validity as it mainly draws on national institutional structures and definitions, which differ significantly from one country to another.

In the recent years, OECD has also attempted to develop a limited set of up-to-date

and internationally comparable indicators on education. There are some problems of relevance and usefulness of the OECD indicators for the developing countries. The advanced countries with their better accounting and reporting systems can produce statistics which are practically impossible to generate for the developing countries. For example, the data on participation in educational programs outside the formal school sector is not reliable in the developing countries. Similarly, the enrolment is underestimated to the extent that private unrecognised institutions contribute to the spread of primary education, skill formation through apprentice training and to literacy through non-formal and adult literacy programs. Some issues in educational expenditure also need stricter scrutiny. How does one estimate and compare the per-pupil cost when the education is highly subsidised in various ways? Finally, many methodological improvements are needed for making educational indicators more meaningful and comparable across national and international systems. It has been found necessary to improve existing set of educational indicators by placing emphasis on educational attainment, literacy and educational flows, and to develop new indicators which can be readily understood and interpreted and which enable direct country comparisons possible.

The Government of India has recently reviewed the progress of DISE implementation and taken two important decisions. First, DISE data collection system would be extended to the whole country by 2003 and secondly, the DISE data would be treated as official data.

The above is only illustrative of the complex task of design and implementation of information system for education which is consistent and globally comparable. The problems are equally serious at the national level. Differences in structure of the educational ladder and varying age of entry to grade I, variations in the entry-level qualifications for teachers, varying degree of subsidisation of educational inputs like books, stationery, uniforms and other inputs pose problems of comparability of the same indicator across states. What type of educational expenditure is included in training and capacity building? How to analyse the economies of scale? What should be the date of reference for recording educational statistics when academic sessions begin at different points of time in a calendar year? How do we decide that a child is enrolled? How to distinguish between attendance and enrolment between working days and instructional time in the educational institutions? How to rank educational institutions in terms of their performance and cost-effectiveness? Is it possible to develop an index of educational development of a region? What is the relationship between the stock and flow of educated manpower?

A national approach on these and other aspects is necessary before the national and international comparisons become meaningful. It is unfortunate that these issues have not attracted the attention of policy planners and administrators. A proper mechanism should be evolved to ensure that definitions and concepts used in education are meaningful and standardised to the extent possible. In order to improve the quality and effectiveness of educational statistics, the field staff engaged in data collection and reporting should be fully trained in the use of definitions and concepts. The following table explains various dimensions of data requirements in educational planning,

decision making and monitoring of various reform programs.

Table 1: EMIS scope and coverage

Spatial unit	Elementary education	Non-formal and adult literacy	Secondary and Senior Secondary	Higher and university education	Institutes of national importance
National	*	*	*	*	*
State	*	*	*	*	*
Regional (NSS type)	*	*	*	*	*
District	*	*	*		
Block/Mandal/Taluk	*	*			
Cluster	*	*			
Institution	*	*	*	*	*

In the Indian context, the district constitutes an important administrative level and is directly related to the operations and maintenance of the school education. It is at the district level that most of data are received and aggregated. The district education officer guides the development of school education in the district and is assisted by a number of block level education officer and subject experts. The district office is also the custodian of all educational data and provides a link between the institutions and the state administration. Each district is also provided with a skeleton staff for data collection, analysis and tabulation of statistical data.

It is also recognised that Indians states are diverse not only in terms of their size but also in terms of social and economic development. Many indices presented at the state level are useful only for administrative and accounting purposes but are inadequate to meet the requirements of planning and management. It is therefore important to consider the concept of a region, which could be smaller than a state and larger than a district. In this context, the NSS regions present a classification which could be adopted for various purposes as the NSS regions happen to be internally more homogenous. With the available technologies, it is possible to develop and maintain databases in terms of regional aggregations.

#### 4. Database on elementary education in India

The education system in India has not only expanded its outreach and coverage but has diversified into new areas with the result that the workload of data collection, collation, tabulation and publication has increased tremendously during the last four decades of development planning. The present day educational system in India comprises of about 600,000 primary, 150,000 upper primary, 70,000 secondary and higher secondary schools. The network of institutions of higher education includes more than 7,000 colleges of general education, about 1000 professional colleges and about 150 universities and a number of specialised institutions of industrial, scientific, technical and social sciences research. These institutions taken together had an enrolment of more than 150 million students and employ about 4 million teachers. As mentioned earlier, the immensity of the data requirements for such a large system is compounded by the multi-level planning and variegated system of educational

administration followed in each state/UT. Developing an integrated data base management system for a widely dispersed network of educational institutions poses enormous challenges to specialised agencies engaged in this task.

#### ***4.1 Types of educational data***

A variety of statistical data is collected from recognised schools in order to meet the needs of educational management at various levels. Generally the data can be classified as follows:

***Institutions:*** Year of establishment, level, rural-urban location, category of institution i.e. for boys only, girls only or co-educational, management, classes taught, VEC/PTA formation and meetings, inspection, number of classes/sections, type of course and specialisation offered especially at secondary level of education etc.

***Enrolment:*** Distribution of enrolment by age, gender and social classification including the data on SC, ST and OBC enrolment, distribution of students by age and class (age-grade matrix), number of sections in each class, new admissions, repeaters and students getting incentives classified by class attending.

***Teaching and other Staff:*** number of teaching and non-teaching staff, their qualifications, classification by gender and caste, trained/untrained, year of joining the service, date of superannuating, place and date of last posting, female principals/head teachers.

***School buildings:*** Type of building, building status (government/private/hired), condition of classrooms, maintenance of school buildings, classrooms requiring repairs and rehabilitation, toilets (general and separately for girls) and their condition, availability of electricity and most frequently used source of water supply.

***Teaching learning materials, Equipment and supplies:*** teachers guides, library books, school bell, storage place, blackboards, chalks and duster, science and maths kit, globes, mats, chairs and tables for teachers, musical instruments etc.

***Education budgets:*** Plan and non-plan expenditure, recurring expenditure, education budget as a proportion of total budget, intra-sectoral allocations for various sub-sectors of education (at constant and current prices). Data on education budgets and its distribution by items is available up to the state level only. District and lower level estimates are extremely difficult to get. Most of the schools are covered by various incentive schemes like free uniforms, free textbooks for selected students, mid-day meals, attendance scholarships etc. The data on these schemes is collected from the department of education and not from the schools. Data on private costs of education at various levels is scarce.

***Income and expenditure:*** receipts, endowments, contingencies, welfare funds collected from students, funds utilised, opening balance and balance carried forward. Schools in DPEP districts receive an additional grant of Rs. 2000 per year and each teacher receives a grant of Rs. 500 per year for preparing instructional materials and other inputs necessary for improving classroom interaction.

***School efficiency:*** teacher pupil ratio, students per class, classroom: teacher ratio,



Intake rate, grade transition rate, drop-out rate, repetition rate, input-output ratio, average years taken to complete primary schooling and other levels by gender and separately for various social groups;

*Learner's attainment:* by subject (mathematics and language), gender, social group, rural/urban areas etc. Share of children able to achieve the desired level of competencies, share of students achieving below threshold competencies. Such data is available for primary education and for the DPEP districts. The NCERT has recently completed the assessment studies at the inception, mid point and terminal stage of the DPEP Phase I districts. There are no assessment studies at the secondary and senior secondary levels.

#### **4.2 Type of allied data**

Educational planners often require a considerable amount of data relating to other sectors of the economy. These requirements will vary depending upon the purpose of inquiry/analysis. However, the following type of data is most commonly needed for educational planning:

*Population:* age distribution by single year, rural-urban distribution of school going age population, structure of population by gender and social classification, age specific population projections for the next 10-15 years;

*Habitation and settlement structure:* In order to examine questions related to access, the population data needs to be supplemented by the geographical data on the location, size and structure of the settlements in a given area. The policy decisions/norms for the provision of schooling facilities are related to the structure of settlement structure. This information is very essential for micro-planning and program formulation in the area specific context. The state governments have prescribed norms for opening of formal school, nonformal centres and centres of alternative schooling.

*Health and Nutrition:* Child survival and development; access to safe drinking water, primary health and nutrition; life expectancy; maternal mortality rate; infant mortality rate, under five mortality rate; incidence of common water borne and infectious diseases; availability of medical facilities.

*Social and economic development indicators:* per capita income, state domestic product, workforce participation and their classification into primary, secondary and tertiary activities. Major agricultural/industrial characteristics of the region, role of NGOs and community in the educational development of the region.

#### **5. Sources of educational and allied data on elementary education: India/states**

At the state level, the Directorate(s) of Education are responsible for the collection, collation, tabulation, analysis and transmission of educational statistics pertaining to institutions, teachers, enrolment, retention, transition and achievement of students in the educational ladder. Data on educational finances is also collected to examine the resource allocation and utilisation patterns. To perform these specialised functions, each state has a planning and statistics cell in the state directorate of education. The

existing system of data reporting has survived since the planning was never taken as a serious empirical exercise and targets were set more on moral and statutory considerations rather than basing them on the strength of analysis of prevailing forces of demand and supply.

The sources of education and allied data at various administrative levels are as under:

- (i) Planning, Monitoring and Statistics Division of the Education Department of the Ministry of Human Resource Development (national level) and the state departments of education;
- (ii) National Council of Educational Research and Training (NCERT) at the institutional, district, state and the national level;
- (iii) NIEPA is responsible for conducting All India Educational Administration Surveys. The last survey was conducted in 1990-91.
- (iv) the Census of India provides data from village upwards at various levels of aggregation on literacy, educational attainment, age distribution of population, facilities, health and family welfare status of the respondent families;
- (v) the National Sample Survey Organisation provide data at the national and the regional level;
- (vi) The University Grants Commission (UGC) collects data on the institutions of higher education.
- (vii) Institute of Applied Manpower Research conducts national surveys on area manpower profiles, vocational and technical education.
- (viii) Indian Council of Medical Research collects data for the medical education in India.
- (ix) Indian Council of Agricultural Research collects data on agricultural education in India.
- (x) Director General of Health Services, Ministry of Health and Family Welfare collects data on health manpower and vital statistics in the country.

In addition to the above sources, many other departments and agencies of the GOI and the state governments collect useful data on education related parameters of the households. A number of university departments, research organisations and development planners and researchers are engaged in collection and dissemination of educational data. The institutions (schools and colleges) maintain their own databases for institutional planning and decision making. In the recent years, many state governments have collected considerable data on household characteristics through door to door survey. The household data collected by the education department is being used to operationalise school mapping and micro-planning at the cluster and the village level. Some international agencies have also sponsored projects and national surveys which provided some very valuable data. Notable among these is the National

Family Health Survey conducted in 1992-93.

### ***5.1 Frequency of data collection and reporting***

The numerical data on education is collected on the basis of academic session whereas the financial data pertains to the fiscal year (April-March). Most of educational data are collected on annual basis with 30<sup>th</sup> September of each year as the record date. As will be observed later some national level data are collected with erratic frequency. The educational statistics of the MHRD are collected on complete enumeration basis. NSS uses sample survey techniques extensively for estimation of socio-economic and educational indicators. Data collection related to education sector was undertaken in 38<sup>th</sup> round (1983-84), 42<sup>nd</sup> round (1987-88), 45<sup>th</sup> round (1990-91) and 52<sup>nd</sup> round in 1995-96 of NSSO. The survey of educational administration conducted by NIEPA was repeated after a lapse of more than 15 years. The DISE data collection system working in DPEP districts collects the data annually with reference to 30<sup>th</sup> September as the record date.

### ***5.2 Data flows and aggregation***

Educational data are regularly collected from all recognised educational institutions for a variety of administrative and planning purposes. For the purpose of educational statistics, unrecognised institutions are excluded for enumeration purposes. Each recognised institution is required to maintain a number of registers which form the backbone of all types of educational statistics.

The primary data collected from educational institutions is aggregated at the block level but the reporting is done at the district, state and the national level. The nature, scope and the coverage of educational data available at each level differs. At the national level, data on a number of variables is pooled to monitor the progress of elementary education. A detailed description of various systems of educational data is presented in the following section.

### ***5.3 National data system at MHRD***

Since educational planning and administration for elementary education is a state subject, the primary responsibility for collection, collation tabulation, analysis and transmission of educational data therefore lies with education departments of the states/UTs. The state/UT level data is aggregated and reported at the national level in the prescribed formats. Collection, compilation, analysis and publication of the national data is co-ordinated by the Department of Education, Ministry of HRD. Data on educational institutions, enrolment, teachers and expenditure are provided by the national system and are collected annually with 30<sup>th</sup> September as the record date. Financial data pertains to 31<sup>st</sup> March of every year. At the national level, efforts to monitor the progress of elementary education in various districts/blocks have not succeeded so far. District level data is available at the state level only. The state do not maintain disaggregated data files.

The MHRD also compiles data on education budgets of the states/UTs. These data are essentially taken from the budget documents of the states/central government.

The MHRD also initiated a number of programs to strengthen the educational data

base. It has also sponsored number of projects to reform the existing system of educational data collection, compilation and analysis. A project on computerisation of educational data, monitoring of enrolment and attendance at the block level and other innovations have failed to take off beyond the experimental stage. A scheme for the strengthening of educational statistics at the state level was launched during the eighth five year plan but had to be discontinued due to poor management and lack of co-ordination among the implementing agencies.

The Ministry of HRD publishes annually two series of data tables. These are 'Selected Educational Statistics' (SES), which contains statewide and levelwise enrolment and the number of teachers by stages. Besides some budgetary data on plan and non-plan expenditure of the states is also given. Estimates of 6-11 and 11-14 age group children are also provided in SES and the GER at primary and middle stage are also calculated. Most of the data in SES is provisional and is finalised with a gap of couple of years. The second publication entitled "Education in India" contains a more detailed and disaggregated data classified by management and classwise enrolment. A separate section provides data for rural institutions. The utility of the data is reduced due to considerable time lags. Further disaggregation of data (Districtwise and below) is not available through these documents. Each level has desegregation up to the level below. The raw data is available only at the district/block level.

#### ***5.4 All India Educational Survey***

The All India educational surveys (AIES) were launched to provide additional information on accessibility (by habitations) and the availability of various types of facilities in schools. The first such survey was conducted in 1957 and provided valuable input for third five year plan when the expansion of educational facilities was taken up on large scale. Subsequent surveys were designed to provide similar information for the subsequent five year plans, but the time delays and erratic periodicity of the surveys has reduced its utility for educational planning purposes. Since the data tabulation and processing is manual, the surveys were not in a position to promote the use of data for micro level planning. Although district level tables are generally available in the NCERT the work of manual compilation is too heavy to be completed before the next survey comes around and the aggregates other than states are difficult to get. The raw data is generally lost, or put to use only by researchers or scholars after a great lapse of time. NIEPA has brought a number of publications based on AIES data, but with considerable delay.

The NCERT conducts AIESs with the assistance of state education departments. There is no permanent structure to collect and analyse the AIES results. The state units which are set up for this purpose are disbanded after the data collection work is completed. The sixth survey was launched in 1993. Preliminary results became available in late 1995 i.e. with a lag of nearly two years. The state specific tabulations have been published recently. Detailed district level data tabulation on access and related aspects are yet not available and may take much longer time. Some states have published state reports whereas in most other states, the detailed data is yet not available to users within the education and outside the education department.

The sixth All India Education Survey is unique in certain respects. The survey was conducted after a gap of eight years as compared to a normal five year period. It was for the first time that NCERT collaborated with NIC for the computerisation of sixth survey data so that the availability improves and data results are available in time. It was envisaged that computerised data will be available at the school, block, district, state and the national level. For those interested in district planning, the block level data will be very useful. The computerisation of AIES has led to considerable problems due to inadequate capacity at the NCERT and lack of co-ordination between the NIC, NCERT and the state survey officers.

It is difficult to comment on the quality of data generated through AIES as there is neither a comparative database nor post-enumeration checks were conducted to establish the quality of data. The MHRD did not collect its statistical returns for 1993-94 as it decided to use the data generated by AIES. It is reported that the quality of AIES is not uniform across states. The data through DPEP (DISE) is also not available for 1993-94.

### *5.5 Census of India*

The population census is the most comprehensive source of data on the educational attainment of the whole population and the workforce. The unit of observation is the household and not the educational institution. The significance of the population data lies in its complete enumeration of the household members along with a large variety of data on social and economic variables. These data are available at various levels of aggregation/disaggregation. The census data reflects the stock of educated manpower and cumulative outcome of the past educational efforts. The data on literacy is available for all levels of spatial aggregation classified by gender and caste. The 1981 census provided for the first time, on a 5 percent sample basis, single year age tabulations for children attending school. The data was available in 1986 but was overlooked by administrators and policy planners. Similar data from 1991 Census has now become available. Its use by the department of education has been minimal, not even for comparative purposes.

The census data can also be used to assess the social demand for elementary education. The population structure and its distribution in school going age group are most commonly used for estimating the demand for education. The census also provides data upto village level on the infrastructure facilities like availability of drinking water, post-office, bank, primary health centre, primary school, non-formal education centre and other amenities.

### *5.6 National Sample Survey (NSS)*

The NSS collects information from selected households on a sample basis. Each round of NSS focuses on specific areas of social and economic activity. 35th round of NSS (1981/82) covered education but the results were not published. A similar inquiry was repeated during 42nd round (1986/87). Preliminary results were available in 1991. A unique feature of NSS data is that it classifies the respondents' educational and other characteristics by social and economic status. For example, the distribution of households by their educational attainment cross-classified by income fractile group is

available from NSS data alone. The NSS in its 35<sup>th</sup> round provided data on children attending schools and the workforce participation rate for those not attending school. Some estimates of the private cost of education were also generated through NSS. This type of data is generally not available from other sources. Recently, NSS has released the preliminary results of the educational data collected through 52<sup>nd</sup> round. Although considerable amount of new information/analysis has been given by the NSSO, the vast amount of data that was collected has not been fully exploited to understand various type of interrelationships. A major reform that is needed in educational and social statistics relates to its availability for social research. NSS data is still not available in full form to the researchers.

The NSS publications contain the following type of educational data:

- Distribution of population by literacy
- Educational attainment of the population
- Educational attainment of the workforce
- Work participation rates of the educated
- Employment/unemployment status of the educated persons
- Children attending schools
- Net enrolment rates
- Per unit household expenditure on various levels of education.
- Estimation of ever enrolled children
- Reasons for drop-out
- Average private expenditure at various levels of education
- Distribution of important educational indicators by different facilities of population.

The validity of the educational data collected by the NSS has yet to be established through independent surveys. Large differences in the estimates of NSS and MHRD data are observed. One does not know how much of it can be attributed to sampling errors and how much to definitional issues. One reason for the differences could be due to different response patterns of the households and the schools to the education related questions. Differences can also emerge due to different definitions and concepts followed. The training of field staff in data collection especially using the educational terminology is also important attribute affecting the quality of data. A unique characteristics of the educational statistics is that the enumerator is the teacher who is also a stakeholder in the educational processes. The NSSO uses the enumerators who are external to the education department. To what extent these variations can be attributed to the bias of the enumerators?

### ***5.7 Planning Commission***

The Planning Commission (PC) monitors the progress of the implementation of the Minimum Needs Program (MNP). Since universal elementary education is one of the action points under MNP, both physical and financial targets are constantly monitored by the Planning Commission. Enrolment of general, SC and ST students by gender and rural/urban origin are major elements of the data generated by the Planning

Commission. In addition data on the release of grants and their utilisation is also available from the Planning Commission. Although, the PC has the most recent data, it differs widely from the data provided by the states to the MHRD. The Planning Commission does not publish educational data but permits users to access it for official purposes only. The differences between the data obtained from the MHRD and the PC are difficult to explain.

### ***5.8 Survey of All India Educational Administration***

There is hardly any data on educational administrators working at various levels. Different states have evolved different structures to suit planning and management of education in their states. It is difficult to assess the relationship between the educational development, management structures and delivery mechanisms including the supervision and inspection system at the state and the district level. Efficiency of educational administration and the type of functions performed by them is difficult to get.

Two surveys on educational administration were conducted by NIEPA. The first survey was conducted about two decades ago when the state level reports were prepared. The second All India Educational Administration survey was conducted in 1992. The detailed reports for about 15 states were published till the end of 1997. The national report and the reports for large states are still awaited. The data for many states has lost its utility due to significant changes in the organisational set up particularly after a number of new programs like DPEP were initiated. These data are now being updated.

### ***5.9 State Level Educational Statistics***

While at the national level, there is a considerable uniformity in the data reporting system, the scope and the formats of data compilation and publication differ widely from state to state. The state data is more detailed as compared to the national data and provide temporal and spatial data on institutions, enrolment, teachers and expenditure (plan and non-plan). While some states provide data on class level enrolment separately for SC, ST, OBC and other students, others do not attach much importance to the publication of educational statistics. Most data is reported by districts. There is a considerable time lag in the publication of educational statistics at the state level. Inadequate and ill-equipped manpower accompanied by difficulties in printing of statistical publications by the government presses cause considerable delays in publication of reports. Some government printing presses are reportedly taking 1-2 years for finalising the publication after the manuscripts are handed over for printing.

The states are also free to collect any additional data through surveys or otherwise on matters having bearing on the policy, planning and implementation of various state level programs. For example, states running mid-day meals scheme might be interested in studying the impact of such schemes on enrolment and retention.

The state legislature has the powers to ask the departments for any data/information needed in the conduct of the house and as may be required by public representatives. Such information is generally compiled from secondary sources of data but sometimes

specific information is also collected from relevant sources.

### **5.10 Other Sources**

As mentioned earlier, there are many other sources that provide educational data. For example, the socio-economic and manpower assessment surveys conducted by the National Council of Applied Economic Research, Institute of Applied Manpower Research, National Institute of Rural Development, Centre for Monitoring of Indian Economy, Central Statistical Organisation etc. provide some very useful data on educational indicators. The university departments of education, SCERTs, DIETs and other teacher training colleges have compiled area specific data.

Other important sources of data include studies conducted for the DPEP and other externally funded projects. Special studies were conducted on tribal education, teacher education, financing of education, social assessment studies and learners achievement studies. The learning achievement study has been repeated after a lapse of three/four years in 42 districts spread over seven states.

## **6. Sources of educational and allied data at the district level**

The nature and type of data available at the district level is somewhat different from that of the national level. The states maintain two types of databases. The first pertains to the reporting arrangements for the national level and the second relates to state specific information required for state level planning and management of education. In this section, the discussion would be confined to both types of databases. At the district level most of the data is not published but is available in the raw as well as in the form of key tabulations. The district educational planners will have to spend considerable time in collecting data from various agencies operating at the district level.

### **6.1 District education office**

The principal source of data at the district level continues to be district education office. All the institutional level data collected from primary and upper primary schools is stored in raw form in the office of DEO. While some data is aggregated by blocks, the disaggregated data may be retrievable at the district level. Data for surveys like All India Education Surveys is also stored at the district level. While computers are available in most districts, the data is tabulated manually. There are no national level guidelines and technical support for computerization of educational statistics. During the Seventh and Eighth Five Year plan period, the central government initiated a centrally sponsored scheme for strengthening of educational statistics. The scheme failed miserably due to unrealistic objectives. The scheme was rejected by the states.

### **6.2 DISE database**

In the districts covered under DPEP, computerised EMIS cells were established at the district level to create and maintain DISE database which includes a time series database on students' enrolment, location characteristics of schools, teachers, school buildings etc. for each primary school in the district. The DISE began with coverage of 42 Phase I districts and provided the base level data for 1995-96. The DISE now covers more than 250 districts covered under DPEP/SSA. The DISE has also been extended to cover schools upto elementary stage of education. From the year 2003, the Government



of India has decided to extent the DISE as the official data reporting system for elementary education. Necessary discussions with the states are undergoing to achieve this objective by the stipulated time. In order to provide technical and professional support, state level EMIS units have already been established in 18 DPEP states. At the national level, NIEPA is responsible for the design of DISE information system and providing the technical and professional support to the states in implementation and maintenance of databases on elementary education. Time series data is available for many years and can be analysed to undertake time series and cross-sectional analysis. The DISE is extensively used by the districts for perspective planning. The Joint Review Missions (DPEP) also use the national report prepared by NIEPA for evaluation of the progress made in access and retention under DPEP interventions.

### ***6.3 School census***

Detailed data on the school going age group population is collected annually through school census in each village. The data is collected from each and every household in the village and all the children in the school going age group are enumerated. The reasons for non-enrollment and discontinuation are also recorded. This data is supposed to be used for enrolling children and also for keeping track of out-of school children. This data largely remains unused.

The schools also maintain record of the average attendance of students. However, this data is available only from the schools. Some summary tables based on average attendance are sent to the district administration for monitoring of school attendance but no system has been developed to systematically analyse and store the attendance data.

### ***6.4 Social and economic data***

The data pertaining to social and economic characteristics is usually available in the office of the District Statistical Officer. The office of the development commissioner/ block development officer also provides data on development activities in the district/state. The District Statistical Officer collects the data from various departments and compiles a handbook of district level statistics for use by government departments.

### ***6.5 District census handbook***

The data on population is available in the District census Handbooks (DCH). The DCH provides a very valuable data on population and the amenities available in each village of the district. It also provides useful data on the topographic features including the transport network in the district. The district and blockwise maps provided in the DCH can form the basis for micro-planning exercises.

### ***6.6 NICNET***

As a part of the national scheme, each district development commissioner has been provided with a NICNET (National Informatics Centre Network) computer terminal. The NICNET centre at the district level maintains data on key indicators for all development sectors and the facility is available to all government departments. NICNET contains data on population, infrastructure, land records, educational

institutions, transport and other sectors of the economy.

### ***6.7 Total literacy campaigns***

In the recent years, considerable data on literacy status of the adult population was collected in districts covered under total literacy campaigns. The data on only provides insight into the magnitude and structure of the illiterates but also throws considerable light on the voluntary agencies, social activists etc. Some evaluation studies were conducted to establish the effectiveness of the TLC approach. Most of the data generated by TLC is not computerised.

### ***6.8 Other sources of data***

In addition to the above there may some specific data available for a particular district. The investigators may therefore try to find out whether the district was a part of any pilot project particularly related to education? Some NGOs have also compiled useful data on social and economic characteristics of the population. Of particular interest may be the data in relation to NFE centres as most of the volunteer agencies are running NFE centres. Sometimes these data could also provide useful information to educational planners.

## **7. Limitations and inadequacies of the data collected**

The review of literature on educational statistics suggests that it suffers from many limitations. These relate to redundancy of data, confusion regarding concepts and definitions, inaccuracies in the reporting and analysis, and irrelevance of data as the emphasis is on the ease of collection rather than on the need of the users. Experience shows that states departments of education have not taken seriously the collection, analysis and use of educational data.

The educational data as collected from educational institutions is primarily used to meet the reporting requirements of the state and the national system. Due to various reasons the actual use of educational data for planning, management, evaluation and monitoring of educational programs is limited. A major constraint on effective educational decision making is the 'information distance' that exists between the educational reality and the decision maker. This distance is best expressed in terms of physical distance, time and availability of information for decision making. Important factors affecting the use of educational data are:

### ***7.1 Conceptual and definitional problems***

There is hardly any standardisation of concepts and definitions that are used in education. While the agencies associated with the MHRD use similar concepts. The reference date for data collection by these agencies is also the same. But the problems regarding the effectiveness of these concepts and definitions remain. Therefore, the definitions and concepts used in educational statistics need refinements and review in the light of experience gained in the last 50 years of educational management in India. For example, when should a child should be considered as enrolled? The practice varies from state to state and leaves enough room for manipulation and distortion of the outcome of the exercise. Should a child who is three year old be enrolled in grade I or not? What is the definition of a repeater? How the dropout rate should be

calculated? There are no statistical manuals/guidelines with the state statistical departments/cells on the standard terms and their definitions. It is high time that necessary guidelines are made available so that consistency could be maintained at the data collection and analysis stage.

While it is true that the number of children getting free textbooks and other incentives can be recorded properly but that does not mean that all the children received full compliment of benefits. These are some issues on which standardisation is needed and corresponding instructions issued so that there is no confusion among the suppliers of data.

The problems become serious when the household data generated by Census and NSSO is compared with similar data generated by MHRD. Since these organisations collect the data from the households, the concepts, coverage and scope of the variables are different. No efforts have been made to integrate the household and institutional data. This will be possible if the focus shifts from target groups to area oriented data collection. The GOI should define various administrative levels for data collection and analysis. All the departments should follow the same codes and levels for data generation. It would therefore be possible to share the data from different organisations like Census, education, health and social welfare departments.

## ***7.2 Useful data wasted***

The existing manual system of data collection, analysis and presentation is highly labour intensive. All operations are carried out manually with limited staff members who are often assigned duties other than dealing with educational statistics. This is a major problem facing the educational statistics cells in the states and the districts.

A perusal of the proforma used for collection of educational statistics will indicate that detailed information is collected on many characteristics. However at various stages of aggregation, vital information is either ignored or is presented in such a manner that it loses its significance. It has been calculated that S-I contains about 250 variables, of which aggregations for only 20 are available at the national level (Selected Educational Statistics). Districtwise and lower level data can not be accessed at the national level. Despite massive efforts that go in for the collection, collation and analysis of data, some basic and simple tabulations like the distribution of schools by enrolment size is not available at all. It is possible to prepare separate tabulations for key data classified by rural-urban, management, school size, distribution of enrolment and teachers by caste, but this is not done. For improving the utility of institutional data, it is necessary that basic data should be available in the disaggregated form at the district and state level for analysis and decision making.

The educational administrators often complain that the available data can not be used for decision making because it is not available in the required and usable form. Most of the decisions thus tend to be ad-hoc or based on partial set of data which may be outdated. Consequently, the returns from investment in data collection are never fully realised.

### *7.3 Lack of co-ordination among data collection agencies*

The educational institutions fall under the management of various departments. For example, in a typical state, the primary schools are run by the department of education, department of tribal welfare, department of social welfare and agencies of the central government like KVS and Navodaya Vidyalaya Society. The data collection is the responsibility of the departments of education who can not exercise any control on the schools run by other agencies/departments. This results in a considerable delay in obtaining data from schools run by other departments/agencies. The non-availability of data from one institution delays the aggregation of blockwise data which in turn causes a cascading effect on higher levels of aggregation. There is no legal basis for co-ordination of educational statistics.

### *7.4 Data gaps*

Although many high power committees have gone into the question of what data should be collected and what should be the data management and reporting mechanisms. The quantum of vital information collected from educational institutions has declined considerably over the years, perhaps with a view to cut short time lags and make the data elements more relevant for planning and decision making. For example, age data in respect of students was collected till early 1970s, when a decision was taken to discontinue it. The MHRD is again considering the introduction of the same realising the significance of age-grade data in planning for universalisation of elementary education. Some other problems of data gaps are:

*Out-of-school children:* It is important to understand the magnitude of the problem of out-of school children. While some data is available from various sources including the National Sample Surveys, Census of India, Department of Education of the central and state governments and the household surveys and other research studies conducted under the DPEP, the disaggregation of children into drop-out and never enrolled is not available. This is very important information for planning of intervention strategies. Some attempts have been made by the DPEP states to collect and analyse the out-of school data. The unit of data collection is household.

The review of the efforts of the states shows some major limitations of this exercise to identify out-of school children. First, the coverage is never universal and the correction factor is not possible due to lack of detailed information on un-enumerated houses; Second, the quality of data varies from one village to another and hence generalizations are not possible; Third, the states lacked expertise and capacity to organize large scale data collection from the households; Fourth, the coverage in urban areas was uncertain. In many states, the urban areas were altogether left out; Fifth, lack of standardization and inadequate training to field staff resulted in poor quality of data reporting.

A fresh effort is made by the Technical Support Group (DPEP) to develop guidelines to assist the states in household survey to estimate the dropout and never enrolled children. The development of habitation level plans, as required under SSA, would also involve the use of household data.

*Participation in primary education:* The number of students attending various type of

schools is not fully known with the result that GER and NER do not reflect the true picture and are sometimes quite misleading. The educational statistics of the MHRD as well as EMIS (DPEP) captures the enrolment in recognized government, local body and aided and unaided schools. However, in the last few years, a number of private unrecognized schools have come up in various parts of the country including a large scale proliferation in rural and educationally backward areas. In addition, other type of facilities like NFE, Alternative schools under DPEP, EGS centres and similar structures have been created through centrally/state sponsored schemes. The management structures, enrolment distribution and other characteristics of such institutions are not known as the schemes have failed to evolve a sound reporting and monitoring system. Therefore, it has become essential to ascertain the contribution of such institutions.

*Unrecognised educational institutions:* Over the years, a large segment of unorganised sector has emerged and encompasses all level of education i.e. from pre-primary to secondary schools. For example, most of the pre-primary facilities are in the private sector for which no data is available. Even at the primary stage, the number of children attending unrecognized schools is large and varies between 5-40 percent. Studies undertaken by the author indicated that in the four DPEP Phase I districts, about 25 percent children were attending primary unrecognized schools.

*Production cum training centres in HRD:* Over the years, the role of on the job education and training is picking up through the use of training institutes and also through distance education mode. While some data is available on distance education, not much is known about the training undergone by individuals.

*Teachers database:* Continuous and comprehensive teacher training is one of the important strategy under DPEP. While an impact assessment of the teacher training is a complex research study, the present study will address some of these questions by ascertaining the views of the VEC members and parents of the learners. It is important to develop a teachers database which could be used for personnel management, grievance redressal, posting and deployment of teachers and for training needs assessment etc.

*Lack of flow statistics:* An assessment of the internal efficiency of the educational system is necessary to optimise the available resources. Calculation of the indicators of internal efficiency requires flow statistics of students. Cohort analysis is one of the methods used for this purpose, where the progress of a cohort of students is monitored over the years. Trends relating to enrolment/ transition/ repetition etc. can be traced very broadly at the national and the state level. This is mainly because the time series data in the form of true cohorts is not available. However, sample cohorts can be studied to estimate efficiency of schools operating under different conditions. There is a need to standardise the methodology for such studies.

*Lack of age-grade matrix:* Enrolment figures are collected without reference to the age of the child in the annual system. Age details were collected in the IVth/Vth AIES on the basis of which net enrolment ratios were estimated. However, available data is not sufficient to indicate trends. While complete enumeration of age-grade matrix may not be possible, efforts to develop sample survey techniques for estimation of age-grade

matrix for various types of schools should be developed on priority basis.

*Attainment levels:* Collection of regular data on achievement is neither a part of the national nor the state system<sup>1</sup>. Recent studies indicate that even when some children manage to reach grade IV/V, they learn a little and their achievement levels are extremely low. However, there is a need to standardise the sampling methodology and conduct sample studies at regular intervals so that progress towards qualitative improvements can be studied. Most of the achievement studies in the recent past pertain to DPEP districts. For some districts comparable data on learners achievement is available at two points of time.

*Incentives:* In order to promote primary education among the girls, SCs and STs, the central and the state governments have introduced a variety of incentives for primary school students. Both pre-matric and post-matric scholarships are available to students belonging to SC and ST students. The states are thus spending a considerable share of their educational budgets on incentives. The incentives include free textbooks, stationery, transport, uniforms, attendance scholarships and school nutrition programs. In fact, the coverage and scope of incentives is increasing as efforts towards universalisation of elementary education are accelerated and so are the budgetary allocations in the state budgets. Neither the central nor the state governments have attempted to collect systematic data on the provision, adequate availability and utilisation of each type of facility and incentives by the beneficiary students. In order to examine the relevance of incentives and to improve their effectiveness and efficiency, regular data needs to be collected and analysed through research studies. This will ensure that the scarce resources are properly targeted.

*Utilisation of facilities:* Ever since the beginning of development planning in India, a number of additional facilities, teaching staff provisions, additional equipment, teaching learning materials were provided to primary and upper primary schools so that quality of learning could be improved. The recently introduced scheme of Operation Blackboard provided financial assistance for the conversion of single teacher school into two teachers with expanded facilities and supply of instructional materials. While the data on the availability and non-availability of facilities and equipment in the schools is collected by the All India Educational Surveys, there is hardly any data to indicate the utilisation of these facilities and effect of the availability of instructional materials on actual classroom interaction and learning outcomes. There are a number of audit reports by the Auditor and Comptroller General of India, which have highlighted the inadequacies of supply condition and utilisation of the materials supplied under various schemes. However, these are more from the financial auditing point of view rather than that of pedagogical considerations. Data on the distribution of institutions by capacity, size, space, daily attendance is also not available for planning purposes. The average number of days for which the school functions in an academic year is an important indicator but data is missing even in AIES data. Actual

---

<sup>1</sup> The analysis of learning attainment of the class IV/V students was systematically taken up in early nineties with the introduction of the concept of Minimum Levels of Learning (MLL). Base line assessment surveys have been carried out in eight to ten states covering about 50 districts.

teaching hours in an academic session are extremely difficult to get. Sample studies should invariably be conducted to assess the impact and utilisation of the facilities created/added in primary and upper primary schools.

*Maintenance and operations of school infrastructure:* The schools buildings have an average of 40-60 years of life depending upon the geographical conditions and the quality of construction. Therefore, periodic maintenance and upkeep is necessary to optimise their life cycle. Studies have shown that if maintenance is not given adequate importance, the life cycle of the buildings gets reduced considerably. It is generally observed that the maintenance and operations of the school buildings and other capital assets, particularly in educationally backward areas, has suffered due to inadequate attention and lack of resources for their upkeep. Practically no data is available to indicate the status of buildings and the financial allocations for the maintenance of school buildings are residual and totally ad-hoc. Under civil construction program of DPEP, such information is now being collected for DPEP districts.

*Non formal education:* Figures quoted always relate to the sanctioned centres as approved by the MHRD. No enrolment and attainment data is available for the NFE centres. There are more than 250,000 NFE centres operating in various parts of the country. In order to put the scheme on firm footing, many states established separate directorate for the NFE. However, this did not help much and the scheme continues to be characterised by many deficiencies although it has been in operation for nearly 15 years now. Despite this the number of out-of school (drop-outs and non-enrolled) is significantly large. Studies for assessing the impact of such programs should be taken on regular basis. A systematic scheme for monitoring of NFE also needs to evolved so that timely interventions are possible. At the district level, there is a need to have better reporting and monitoring system for the NFE program.

*Early childhood education:* The data on early childhood is extremely inadequate. The ECE activities fall within the preview of a number of agencies including the ICDS centres managed by the women and child welfare department. In addition to government departments, the pre-primary education is imparted by the unorganised sector for which no statistics are collected.

*Financial statistics:* India is one of those countries where worthwhile financial statistics are highly inadequate. In the existing system, the educational expenditure is collected by the type of educational institutions rather than by the type of courses of study offered. It is unfortunate that the expenditure on primary and upper primary education is not available separately. The unit cost estimates of primary and upper primary education are not available. The norms for provision of teachers and other facilities are entirely different for primary and upper primary schools.

In addition to the public expenditure on education, the private expenditure also constitutes a significant part. There are no estimates of the private (household) expenditure on education. Financial data on the following aspects will be vitally needed for analysing the internal and external efficiency of school education:

- Expenditure by level and course of study.
- Unit cost for various level of education.

- Share of expenditure on education as a proportion of SDP/GDP.
- Share of expenditure on education by sources.
- Itemwise expenditure on various levels/course of study including the cost of direct and indirect subsidies.

### *7.5 Time lag*

The practical utility of whatever data is being collected, collated and tabulated is further reduced due to large time lags in its availability. It takes anywhere 3-5 years to finalise the national data for all states. Districtwise data, originally meant to be collected quinquennially at the national level, is now 10-15 years old, and is not even available for all the districts for the mid eighties. The problems were so immense that this compilation at the national level has now been discontinued.

In order to reduce time lags, the MHRD introduced a shorter version of 'Education in India' in the form of 'Selected Educational Statistics' (SES). The data for SES is collected by canvassing a separate 'Summary Sheet' alongwith the main forms. The summary sheet has basic numerical data on gross enrolment at each stage, number of institutions as well as number of teachers. This stratagem has considerably improved timeliness, but in doing so has (i) further increased time lags with the main formats and (ii) sacrificed reliability, by giving figures relating to previous years for those states that do not submit information in time and (iii) reconciliation problems between the summary and final data on educational indicators.

The MHRD had also tried computerisation of data at the national level for the two publications but with little improvement in time lags. Other experiments promoted by the central government have also made little progress due to lack of interest among the state administration. Large scale computerisation is suffering due to lack of adequate financial resources and limited availability of professional manpower to implement the computerisation on a large scale.

### *7.6 Reliability and redundancy of data*

The issues related to the reliability and quality of educational data have persisted for a long period. The problem is more serious for enrolment data. The quality of data is affected by several factors:

- (i) the practice of over reporting enrollment to show that targets have been fulfilled and also to justify the additional requirements for teachers;
- (ii) the data enumeration and tabulation errors due to manual handling of huge quantities of data;
- (iii) little use of data in planning, management or day-to-day administration;
- (iv) low priority for systematisation of educational statistics; and
- (v) no cross-checking and validation of data.

While the multiple sources of data on the selected indicators can be effectively used as validation checks, it is not being done at the moment, No independent post-



enumeration checks are carried out to establish the validity of educational data. It is necessary to promote the use of sample studies for estimation of key indicators of school performance.

### ***7.7 Developing comparable time series data***

While the time series data is easily available at the national and the state level, no efforts are made to generate and disseminate similar data for the district and the sub-district level. Reorganisation of districts/blocks as well as changes in the concepts and definitions create confusion in the interpretation of data emanating from various sources. Systematic efforts should be made to develop districtwise time series data on selected indicators of educational development. The national/state level resource institutions can assist by taking case studies in selected states.

### ***7.8 Linking data across populations and levels***

Due to high level of aggregations, it is not possible to examine all data across all relevant levels and sub-populations nor to link all data across all levels and sub-population groups. For example, it is not possible to relate teachers data with the schools where they are employed. This factor limits the use of educational data for analysis of inter-relationships between various variables. This is one of the major limitations of the existing data that can be overcome only by computerisation of institutional level data.

### ***7.9 Weak capacity to handle educational statistics***

The present administrative arrangements at the centre and the state level are highly inadequate to meet the growing needs of data management. Lack of adequate investment in modernisation is another related issue. Most of the states either do not have any staff exclusively for data management or the staff is being used for other administrative work.

## **8. Recent attempts to revitalize educational statistics**

Data needs for planning and management of education have increased considerably with a greater emphasis on micro-level planning and the involvement of NGOs, VECs and other voluntary organisations in the implementation of the programs of UPE. With the setting up of Panchayats in all states, a number of management functions will get transferred to them. Similarly, in the urban areas, local bodies have been assigned the management of school education. For the last few years, the GOI has undertaken a massive program for UPE which covers about 150 educationally backward districts of the country (DPEP). In order to meet the data requirements for planning, management and monitoring functions at various levels, a number of initiatives have been taken in the recent past. These include:

- (a) The MHRD introduced partial computerisation of the national data collection system in 1986. The contents of the data capture formats remained the same but the forms were changed to make computer friendly for data entry purposes. The system required data to be aggregated manually at the block and then passed on directly to the state level to be fed into computers. This system was extended to nine states by 1991. While experience with the pilot project in U.P. was good, the

problems of data validation in many states caused innumerable delays. The progress was quite mixed and delays in manual aggregation and transferring of data from one level to the other persisted. Most states have given up computerisation.

- (b) In order to facilitate the speedier transmission of data using the NICNET services, formats were redesigned to create a much wider data base which also included the type of data collected by the All India Educational Survey. NIC left the option to the state government to accept, adopt or redraft the format. Some States (Maharashtra, Karnataka and J&K) have taken positive action. All of them have begun widening the data base even further, sometimes to an extent where it has become impractical for regular and timely handling. Age specific details of enrolment have not been included by any of them. Despite initial success, the tempo could not be maintained. No state has been able to reach sustainable level. Recently, the Department of Education in Andhra Pradesh has designed a monthly and quarterly reporting system on school attendance, teachers and enrolment. The project is still in testing phase.
- (c) NIEPA and Ministry together started a pilot project (COPE) in 1989 to develop software for computerisation of institutional level data and making it available at the district level so as to help the DEO in his day-to-day management. The same data could be aggregated to feed the state and the national level data bases. Schoolwise data was entered into a separate computer supplied to the DEO. The pilot project initially implemented in seven districts was considered successful, and a decision was taken to start a central sector scheme to phase this system into all the districts within three years and to replace the forms of the existing national data collection system with this database. The COPE system has been implemented in 29 districts of Madhya Pradesh with UNICEF assistance and has been running satisfactorily. A state level centre was also established to provide technical support to the district units. Although a decision was taken to implement the COPE system in all the districts, the progress was hampered as the hardware was not available in all the districts and there was no provision for the appointment of professional staff at the state and the district level.
- (d) With the renewed commitment for UPE, the MHRD initiated a District Primary Education Program, which emphasised a greater role for the district in planning and management of primary education. A number of external agencies also provided technical and financial support for the program which is slated to begin by latter half of 1994. An innovative approach to the development of MIS for the educational statistics and also for the project management is visualised. The UNICEF has supported a pilot project at NIEPA for the development of software for institutional data base and also for the development of project management and monitoring software (District Information System for Education- DISE). The DISE has now been implemented in all DPEP districts (about 250). The pilot project has demonstrated the capabilities of the software and also provides the necessary technical and professional support to the district and the state staff till they become self sufficient. The coverage under DISE has since been extended to upper primary

education.

## 9. DISE database and its features

DISE is now an integral part of the EMIS being implemented in the DPEP districts in India and is operational in about 250 districts spread across 18 states. The first implementation started in 1995-96 and by now time series data for six years has been generated for about 40 districts, two to four years data for about 100 districts and single year data for about 100 districts. The latest data pertaining to September, 2001 is available for all schools in more than 200 districts of the country. The important features of DISE are:

- a) DISE is designed as a national system of information for primary education to be implemented in all DPEP districts. Realising the usefulness of DISE, many non DPEP states have also started data collection using the DISE formats. Upper primary education has also been added from 2001 onwards. As per the recent decision of the MHRD, it would gradually replace the existing manual system in other districts also.
- b) It overcomes most of the constraints of the manual system. Easy to install and maintain. Bridges gaps between the producers and users of data/information.
- c) It is an integrated system and is designed to cut short time delays in the availability of database in the final form. It is expected that the reports would become available within 8-12 weeks of data collection. The districts completed the data entry and analysis much earlier.
- d) The school database contains a minimum amount of data that will be collected from all recognised primary schools/sections covered under DPEP. It is consistent with the national system of educational data for primary education.
- e) DISE is the only source of data on items like:
  - Age grade matrix by caste, gender and class.
  - Information on NER is provided from Block level onwards.
  - Enrolment pyramid (by classes, by SC/ST enrolment)
  - Vital information on access (villages without schools, school size)
  - Indicators of internal efficiency like repetition rate, drop-out rate etc.
  - Type and status of school buildings/classrooms
  - Formal inspection of schools
  - Receipt and expenditure in respect of DPEP contingent fund
  - Classification of schools by selected attributes
  - School working days in a year
  - Establishment and meetings of VECs

- Classification of school data by management, school size, rural/urban character and many other important characteristics
  - School level PTR are compared with the average for Block/Taluk.
  - School summary report to be shared with VECs
  - Block and cluster level reports for analysis of status and program interventions.
  - School level information can be assessed from the district, state and even from the national level.
  - Provides comparable information of key performance indicators *across time and space*. Key indicators include GER, NER, PTR etc.
  - Comparative analysis of DPEP and non-DPEP schools.
  - Distribution of students by medium of instruction.
  - Examination results at the end of primary and upper primary grades.
  - Classwise enrolment of children with disabilities.
  - Students flow information.
  - Database on individual teachers including sex, caste, age, professional and educational qualifications, inservice training and days spent on other activities.
- f) While the district acts as a nodal centre for collection, storage, management and maintenance of primary database, it will be simultaneously shared at the state and the national level.
- g) Special controls and checks have been built in to make sure that duplicate records are not included and non-responding schools are tracked instantly.
- h) The school database will form a common core of comparable data for all the DPEP District. The same formats will be used for the collection of core data.
- i) There is a provision for data entry in multi-user mode so as to reduce the time required for data entry.
- j) The software provides data aggregation at multiple levels. Comparable reporting and analysis is possible at the school, cluster, block/Taluk/Mandal, district, state and the national level. It is thus possible to compare the performance indicators for two schools across different districts/states, a possibility that never existed before.
- k) The software can provide on-line data for a number of years. The software contains provision for the time series analysis of data. It would thus be possible to monitor the trends of key performance indicators at various levels of geographical aggregation.
- l) Selected standard reports generated at the district level will flow from district to the state and the national level. Block and Cluster level reports will also be generated for sharing among the BEO/AEOs and other officials working at the cluster level.

- m) For in-depth analysis and relating it other social and economic information, the DISE data can be exported to other packages like LOTUS, EXCEL, Dbase, SPSS etc.
- n) There is a provision for the bi-directional flow of data. The reverse flow of data is from the district to the Block/Taluk and to the school for verification and sharing among the members of the VEC. the report incorporates general information, enrolment data and a number of key indicators of school performance. Similarly the data can be transferred and shared with higher level of users at the state and the national level.
- o) The District will have the facility to generate additional data elements, if deemed necessary. The additional modules for data analysis can be state level and will not from part of the national database.

In addition a powerful query module assists the educational administrators in generating reports and classification of schools according to multi-level conditional criteria. Many features in DISE are designed to assist the educational administrators in improving the quality of data:

- a) System generated validation, consistency and exception list generation.
- b) Improved transparency and sharing among user agencies.
- c) Quick availability of data will permit post-enumeration checks.
- d) It will be possible to detect defaulting schools and identify the schools with exception behaviour of selected indicators.
- e) Sharing of school summary report with VEC and other community leaders will help in identification of discrepancies in the actual and the reported data.
- f) Increased use and awareness will help in improving the quality of data reporting.
- g) Errors due to manual processing are completely eliminated

Considering the success of the DISE experiment, a decision was taken in 1998 to develop the state and national level systems so that an integrated model of data flows and dissemination could be undertaken. Accordingly, the state level and the national level system of data flows and data management have been developed and are in place. It would thus be possible to monitor the trends in key performance indicators over time and use these for policy planning and review purposes.

## 10. Future possibilities

In order to revitalize the educational statistics, a number of steps will need to be undertaken both at the state as well as the central level. Some of these are discussed below.

*Defining performance indicators:* Little work has been done to evolve performance indicators for the school system. Indicators like GER and teacher:pupil ratio are important but are not sufficient to describe the performance of the school system.

There is a need to develop a set of simple and easily measurable set of performance indicators at the school, district, state and the national level. The progress on all aspects of access, retention and achievement and efficiency should be clearly reflected by these indicators. Once an agreement is reached on these indicators, a system of data collection and monitoring can be developed with the full involvement of the state and the central government(s).

*Coverage:* Presently, the school statistics cover only the recognized educational institutions. These include both aided and unaided schools. However, in the recent years there has been a great proliferation of a large number of unrecognized primary and upper primary schools in both rural as well as urban areas. Thus the actual enrolment for the children attending primary and upper primary is under-estimated. As of now, there are no studies which have estimated the magnitude of enrolment, type and quality of education in the unrecognized private schools. In order to realistically assess the magnitude of non-enrolled children, an adjustment has to be made for the enrolment in unrecognized private schools. Sample studies can be conducted in selected districts to develop a methodology and framework for estimation of enrolment in private unrecognized schools.

*Scope of data collection:* The present system collects data only on the input and output of the school system. There is hardly any data on the process variables. With a great emphasis on the planning and management of quality improvement, the need for data information on process variables has become almost a necessity. These data are presently being collected through ad-hoc studies. Data on the utilization and impact of facilities like blackboards, kits, books, teacher training and special incentives is not being collected at the moment. School attendance and working days is another area on which little is known at the moment. It is possible to collect some vital information through sample studies. A system of 20% sample data collection alongwith the collection of other educational statistics can be introduced for sample studies on resource utilization. However, the precise methodology needs to be worked out.

*Strengthening data collection machinery:* Presently school statistics are collected and processed by the educational administrators who are also engaged in the implementation of various educational programs. In order to show quantitative progress particularly in terms of enrolment, data are fudged at various levels. The existing system of data aggregation precludes the possibility of identifying the source of error. The present system of DISE implemented in DPEP districts is a computerized model and will only be affected by mis-reporting at the school level. The sharing of school summary report with the members of VEC makes the school data transparent and the chances of manipulation will thus be minimized. However, this assumes the existence of vigilant local community and VEC.

*Capacity building in the use of EMIS data:* Due to inordinate delays and problems related to the quality of data, the use of data for actual management, decision making and for monitoring purposes is getting minimized. This trend has to be reversed. A multi-pronged strategy for capacity building of the educational administrators at various levels in the use of computers and computer based information systems will be required. This is particularly important as most of educational administrators are in

late forties and early fifties and have very little exposure to computer applications in educational planning and administration. In order to implement this program, a number of regional and state level resource institutions will have to be developed. The DIETs, SCERTs and other research institutions can play a vital role in capacity building of educational administrators.

*Exploring alternative sources of data:* In addition to regular educational statistics, a number of other agencies generate significant amounts of educational data. These include All India Educational Surveys (NCERT), population census (Registrar General of India), National Sample Surveys and nationwide research studies relating education and allied spheres like health, population, nutrition etc. The All India Educational Surveys, being the most important source of supplementary data, have lost their utility due to bad planning and lack of foresight in planning and management of large scale surveys. The data generated by other sources is hardly used, and not even examined, by the state and central departments of education. Thus some very vital data is not used at all.

*Modernization of the methodologies of data collection:* The existing system of complete enumeration of all schools (recognized) has outlived its utility. There have been significant developments in statistics which have now made it possible to generate estimates of key performance indicators. These techniques have been successfully applied for estimation of foodgrain production in the country for many decades. It is possible to generate estimates of enrolment and other performance indicators by collecting data on sample basis. Moreover, complete enumeration of certain indicators is just not possible. The only alternative in such cases is to generate estimates through sample studies. Gradually the load of regular data collection can be reduced by switching to sample estimates. The present system of data collection under DPEP will provide the basis for the development of sample studies. To begin with, the use of sample surveys can be tried in some selected DPEP districts.

*Post-enumeration checks:* In order to assess the reliability and accuracy of the data collection, a provision of post-enumeration check is always built in. This concept has been fully utilized in population census and other large scale surveys. Unfortunately it has never been practiced in the area of educational statistics. These checks could be carried out by independent agencies. This will be a vital step in building the confidence and improving the quality of educational statistics.

As noted above, implementation of DISE is one of the first large scale effort to computerise the educational statistics at the district level and to build linkages at the state and the national level. DISE experiment has shown initial success and it will be possible to have disaggregated schoolwise and teacherwise data which can be timely, reliable and far more complete and available simultaneously at the district, state and the national level. It would also be possible to change the target setting systems for the state governments and set targets for retention and completion along with enrollment. Alternate scenarios can be developed and their feasibility studied through quick simulation models. Given the propensity of the administration to respond to targets, this would be one of the important ways of shifting state priorities from expansion alone onto qualitative improvement as well. However, a strong commitment to the DISE

model will be required from the DPEP, the central and the state governments.

Over a period of time, DISE will be able to provide vital information on the following:

- regular educational statistics (annual) aggregated at block, district, state and the national level.
- provide quick estimates of indicators like transition rates, drop-out rates and percentage of underage and overage children;
- provide a complete age-grade matrix for each level of aggregation. This data is not available at the moment but is vital for estimation of net enrolment and the coverage of 6 year children.
- provide institutional and related data for micro-level planning and management of primary education;
- classify institutions in terms of their performance over the years and monitor the manner in which this classification changes;
- monitor the implementation of various programs at the district, state and the national level.
- generate a master list of schools and teachers for further research and other studies.

An important issue at this stage is whether the system wide reforms in educational system can be brought about using the DISE experience or not? The answer is yes, as the DISE is a step in the right direction. It can be further extended to cover other levels of education so that an integrated view of the whole system of education, at least covering all stage of school education could be undertaken.

## 11. Summing up

The above discussion while reviewing the need for educational data has highlighted the various sources of data and also pointed to the data gaps. It has also emphasised the need for modernisation of data collection, handling, storage and sharing of key data among a variety of users. There is a need for capacity building among educational administrators so as to promote the use of educational data for planning, management, monitoring, evaluation and decision making. For the success of district planning, it is essential that institutional level database must be generated and maintained atleast at the district level and preferably up to the state level.



## REFERENCES

- Aggarwal, Yash (2002). *Regaining Lost Opportunity: The Malaise of School Inefficiency*. NIEPA, New Delhi.
- Aggarwal Yash (2001). *Progress Towards Universal Access and Retention: Analytical Report*. NIEPA, New Delhi.
- Aggarwal Yash (1988), *Education and Human Resource Development*, Commonwealth, New Delhi.
- Aggarwal Yash (1998a), *Access and Retention: The Impact of DPEP*, MHRD, New Delhi.
- Aggarwal Yash (1998b): *District Information System for Education: User manual*, NIEPA, New Delhi.
- APEID (1985), *Towards equality of Educational Opportunities*, UNESCO, Bangkok.
- Bryk, A.S. & Hermanson, K.L. (1993). Educational Indicator Systems: Observations on Their Structure, Interpretation and Use. *Review of Research in Education*, 19, 451-484.
- Caillods, F. (1991), *Educational Planning for the Year 2000*, IIEP Contributions No. 4, IIEP, Paris.
- Climaco, C. (1992). Getting to Know Schools Using Performance Indicators: Criteria, Indicators and Processes. *Educational Review*, 44, 295-308.
- Haddad W.D. (1995), *Education Policy-Planning Process: An Applied Framework*, UNESCO-IIEP, Paris.
- Haq Mahbub Ul and Khadija Haq (1998), *Human Development in South Asia: 1998*, HDC, Islamabad.
- MHRD (1999). Current Status of Educational Statistics and Emerging Needs from National and International perspective, Department of Statistics, Ministry of Human resource development, New Delhi (mimeo).
- MHRD (1998), *Selected Educational Statistics 1997-98*, MHRD, New Delhi.
- MHRD (1997), *Annual Report 1997*, MHRD, New Delhi.
- Oakes, J. (1986). *Educational Indicators: A Guide for Policymakers*. New Brunswick: New Jersey Center for Policy Research in Education, Rutgers University.
- Oakes, J. (1989). What Educational Indicators? The Case for Assessing the School Context. *Educational Evaluation and Policy Analysis*, 11, 181-199.
- OECD (1997), *Education at a Glance: OECD Indicators, 1997*, Centre for Educational Research and Innovation, OECD, Paris.
- OECD, (1992). *The OECD International Education Indicators: A Framework for Analysis*. OECD, Paris.
- Porter, A.C. (1991). Creating a System of School Process Indicators. *Educational Evaluation and Policy Analysis*, 13(1), 13-29.
- Ross K.N. and Mahlek L. (ed), (1990), *Planning the Quality of Education: The Collection and Use of Data in Informed Decision Making*, UNESCO-IIEP, Paris.
- Tyagi P.N. (1993), *Education for All: A Graphic Presentation*, NIEPA, New Delhi.
- UNDP (1997), *Human Development Report, 1997*, UNDP, New York.
- UNESCO (1993), *World Education Report, 1995*, UNESCO, Paris.

UNESCO (1995), *Statistical Yearbook, 1995*, UNESCO, Paris.

UNICEF (1997), *Child Info: Database and mapping Information System*, UNICEF, Kathmandu.

UNICEF (1997), *Statistics of South Asian Children and Women*, UNICEF Regional Office for South Asia, Kathmandu.

Windham (1988), Effectiveness Indicators in the Analysis of Educational Activities, *International Journal of Educational research*, Vol. 12(6): 575-666.

Windham D. (1988), *Indicators of Educational effectiveness and efficiency*, Tallhassee, Florida.

World Bank (1996), *Trends in Developing Economies*, World Bank, Washington.

World Bank (1997), *World Development Report: 1997*, World Bank, Washington.

-----