

## CHAPTER 3

### DATABASE AND RESEARCH METHODOLOGY

#### 3.1: Sampling Design

Schools are selected by purposive and stratified random sampling from Midnapore town and its surrounding rural areas in the Paschim Medinipur District. For a detailed study, schools are selected by stratified and purposive sampling. Total sixteen primary schools have been selected from the entire population of the schools in Medinipur and its surrounding rural areas in the district of Paschim Medinipur, West Bengal, India. Stratifications are based on the differences in the principles of management of schools (private and public) and the locations of the schools (rural and urban).

#### 3.2. Variables:

*Dependent variables:* students' enrolment and students' learning achievement score. These are quantitative variables.

*Independent variables (for students' enrolment):*

We have considered total 11 variables. These variables are –

1) academic qualification of guardians, 2) average annual household income, 3) school fees, 4) learning achievement score, (5) quality of teaching index, 6) school infrastructure index, 7) location, 8) organization, 9) female-teacher ratio, 10) availability of UP (upper primary)/secondary schools (within 1 k.m. to 2 k.m.) and 11) headmaster as a leader. Among these

variables, location, organization and availability of UP (upper primary)/ secondary schools (within 1 k.m. to 2 k.m.) are qualitative variables and rest of the variables are quantitative variables.

***Independent variables (for students' learning achievement score):***

We have considered total 16 variables. Of them, 15 are qualitative and only one is quantitative. The qualitative variables are -1) academic qualification of parents (father and/or mother), 2) source of income, 3) school infrastructure, 4) private coaching, 5) school organization, 6) location of school, 7) gender of students, 8) caste of the students, 9) school attendance of the students and 10) the quantitative variable is annual income of the household of the students. Some of the qualitative variables are dichotomous (yes or no) and some are trichotomous or more. For example, the variable- academic qualification of the parents has three characteristics such as i) graduates or higher, represented by  $(D_{2i})$ , ii) H.S. /M.P. (Higher Secondary or Madhyamik Pariksha )  $(D_{3i})$  and below secondary which is the reference characteristic. The characteristics of the other dummy or qualitative variables are as given below.

<b>VARIABLES</b>	<b>CHARACTERISTICS</b>
<b>Source of Income</b>	Service(D4), Business(D5) and Family(D6) and others(reference characteristic or variable)
<b>School Infrastructure</b>	Excellent(D7), Very good(VG)(D8), Good(G)(D9) and Not satisfactory ( the reference characteristic or variable)
<b>Private Tuition/Coaching</b>	Yes(D10), No(reference characteristic or variable)
<b>School Organization</b>	Private(D11), Public(Reference Variable)
<b>Location of Schools</b>	Urban(D12), Rural(Reference Variable)
<b>Gender of the Students</b>	Female(D13), Male(Reference Variable)
<b>Caste Of the Students</b>	General Caste(D14), SC,ST & OBC(Reference Variable)
<b>School attendance</b>	Very Good(D15), Good(D16), Not good(Reference Variable)

Schools are selected by stratified purposive and random sampling from Midnapore town and its surrounding rural areas in the Paschim Medinipur District. The number of schools thus selected, is Sixteen (i.e. 16). The full lists of the primary schools selected are given below.

#### **RURAL PUBLIC**

Sl. no.	Name of Schools
S1	PIRAKATA PRIMARY SCHOOL
S2	SATPATI PRIMARY SCHOOL
S3	CHAINPUR DAKSINSOLE PRIMARY SCHOOL
S4	MALIDA PRIMARY SCHOOL

#### **URBAN PUBLIC**

Sl. no.	Name of Schools
S5	SAHID ANATH BANDHU ADIBASI PRIMARY SCHOOL
S6	PAHARIPUR PRIMARY SCHOOL
S7	MIDNAPUR TOWN PRIMARY SCHOOL
S8	PATHARGHATA PRIMARY SCHOOL

### URBAN PRIVATE

Sl. no.	Name of Schools
S9	SHREE SHREE BALANANDA PRIMARY SCHOOL
S10	SARASWATI SHISHU MANDIR (SARATPALLY)
S11	ADARSHA SHISHU BIKASH KENDRA
S12	SARASWATI SHISHU MANDIR (BOXIBAZAR)

### RURAL PRIVATE

Sl. no.	Name of Schools
S13	RAMKRISNA SAISAB S. KENDRA
S14	BHIMPORE V. CHILD ACADEMY
S15	PIRAKATA MODEL SCHOOL
S16	AMBEDKAR A.S. NIKETAN

The methodology adopted for this micro study is primarily Quantitative and Qualitative and interpretative in nature, and it attempts to capture the local dynamics that frame social, economic and gender equity issues in primary schooling in district of Paschim Medinipur in West Bengal.

This study is designed to carry out an in-depth quantitative and qualitative analysis of the educational status and performance of the Primary schools. It focuses primarily on the educational facilities and infrastructure available at the primary level.

DATA COLLECTED ON THE BASIS OF QUESTIONS ON THE ISSUES ON

### **The Different Primary Schooling Institutions and the Extent of Private Schooling**

- ✓ Different primary schools
- ✓ Enrolment in different primary schools
- ✓ Gender and Social categories in different primary schools

### **Objective Conditions of Different Types of Schools**

- ✓ School Infrastructure
- ✓ Teachers
- ✓ Educational qualification and training of teachers
- ✓ Salary of teachers
- ✓ Working days and teaching time in the different primary schools
- ✓ Working hours in a day
- ✓ Subjects taught
- ✓ Methods of teaching
- ✓ Cost of schooling
- ✓ School charges
- ✓ Total cost of schooling
- ✓ Annual income of the household

- ✓ Sources of income of the household
- ✓ Guardian qualifications
- ✓ Management

### **Delivery of Education in the Different types of Primary Schools**

- ✓ Attendance in different types of primary schools
- ✓ Parents' view on the performances of the schools, teachers and children
- ✓ Assistance in learning outside the school
- ✓ Assistance at home
- ✓ Private Tuition
- ✓ Learning achievement of children
- ✓ Correlation of annual Income of the family and learning achievement
- ✓ Parents' participation in the governance of school
- ✓ Parents' visits to school
- ✓ Choice of School

The methodology was decided upon in consultation with some experts and researchers; it sought to combine direct observation, informal conversation, open ended interviews to capture the local dynamics of access, enrolment and retention in primary schooling at the micro level, especially in the last January, 2017 to June, 2018. The primary focus of this study was on the children, teachers, parents and the community (especially its manifestation with regard to school committees) and their experiences of schooling. Open ended interviewing was actively encouraged to explore and understand the attitudes, opinions, feelings and behavior of individuals, especially teachers. Direct observation was most helpful in documenting classroom

processes to amplify the teaching and learning that take place in school and also to flesh out the role of the social identities of teachers and students in framing these interactions.

### **3.3: Research Tools**

The research instruments used for this study are

- DISE Data and Official records
- Semi-structured interview schedules and Interview schedules

These schedules are formed on the basis of **questionnaire** developed on the lines of study, published by Pratiche (India) Trust (2006), and SCERT (2009)

- Leadership Survey(46- Items)[developed by Wilson, Deborah(2013)]

The survey consisted of 46 questions representing independent variables: Monitoring Student performance (survey items 1-5), School Climate (survey items 6-11), Instructional Supervision (survey items 12-16), Leadership Style (survey items 17-20), Structured Operational Environment (survey items 21-25), Teacher Recognition (survey items 26-29), Teacher Training (survey items 30-34), Team Collaboration (survey items 35-38), Shared Leadership and Decision Making (survey items 39-42), and School Encouragement of Parental Engagement (survey items 43-46). The survey items were developed and measured utilizing a Likert-scale with the following response options: 1 = strongly disagree, 2 = disagree, 3 = uncertain, 4 = agree, and 5 = strongly agree. It has been tested, validated, and normed. For example, Cronbach's Alpha coefficients ranging from 0.791 to 0.842 have been reported for ten sub-scales.

All survey questionnaires are attached in the appendices (Appendix-A to Appendix- H)

### 3.4 Collection of data

#### *Sources of Data:*

- **Primary data** were collected for students from sixteen primary schools in the Paschim Medinipur district, through personal interviews and mailing questionnaire to the concerned school teachers, headmaster etc.
- DISE data with the verification of official records of primary school students in sixteen schools in the Paschim Medinipur district served as the **Secondary source of data** for analyzing access to primary education. In addition, data were also collected from District Inspector of Schools' Office/ Sub-Inspector of Schools' Office.

#### **Procedure:**

The participants (Students, Head Master, Teachers, Parents, and School Inspectors of schools) were approached personally and after establishing rapport with them the interview procedure was conducted on a one-to-one basis. During interviews, the participants were probed as and when needed in order to gain an in-depth understanding the different factors responsible for under representation of gender and social groups in primary education.

For student enrollment, one Head Master, 2 teachers, 10 guardians, 10 students of class one to class four from each school and corresponding School Inspector (SI) of schools were asked on the basis of questionnaire and for learning achievement score one Head Master, 2 teachers, 10 guardians, 10 students of class four from each school and corresponding School Inspector (SI) of schools were asked on the basis of questionnaire.



### 3.5: Analysis and Interpretation of data

Percentage, F-test and Modified Sopher's Disparity Index (MSDI), Likert Scale, Quality of Education Index (QEI) and regression analysis etc. are used for quantitative analysis of data. Data analyses were performed with the help of EXCEL, SPSS 23.0 software and E-Views.

In this study, we used the Disparity Index proposed by David Sopher (1974) and altered by Kundu and Rao (1986) and revised further by Kundu (2005). The index quantifies disparity between two groups based on the percentages of their populations in custody of a particular property (in this case, enrolment of primary education) in terms of the logarithm of the odds ratio - that is the ratio of the odds that any member of one group (in location / organization / gender / caste) has enrolled in particular school to the odds that any member of the other group does.

In brief, if X1 and X2 are the percentages of males and females enrolling for primary education respectively, then the disparity index (DI) is given by:

$$DI = \text{Log} ( X2/X1) + \text{Log} \{(100-X1)/ (100-X2)\}.$$

The leveling-off effect (items / attributes with high attainment levels may have relatively low disparity than the attributes with low levels of attainments although there is same gap for both items / attributes) can be reduced by taking log (Sopher,1980).

Kundu (2005) have shown that the above index fails to satisfy the additive monotonicity axiom. They have, therefore, proposed a modification to this Index as follows:

$$MSDI= \text{Log}(X2/X1) + \text{Log} \{(200-X1)/ (200-X2)\}.$$

When Q is less than or equal to 200 and X1 is less than X2; X1 represents the deprived section of a community, i.e., here the female student-teacher ratio and X2 are the residuals, i.e. the male student-teacher ratio. That the value of MSDI is zero implies that there is no inequality between male and female students-teachers ratio.

Now, this index satisfies the following axioms: the axiom of Additive Monotonicity, the axiom of Redistribution, the axiom of Repetitive Transfers and the axiom of Multiplicative Monotonicity [Sen(1973) and Kakwani(1977)].

Based on these above mentioned properties, Modified Sopher's Disparity Index (MSDI) is developed. With this index we examine location-wise, caste-wise, organization-wise, gender-wise variations in the enrolment of the students in Primary education.

Again, Quality of Education Index (QEI) has been constructed for each school. The QEI is a simple average of the three indices, namely, Quality of Teaching Index (QTI), School Infrastructure Index (SII) and Learning Achievement Index (LAI), i.e.

$$QEI = (QTI + SII + LAI) / 3$$

Thus, the QEI will indicate the quality of primary education in the schools under study. It is worth mentioning here that all of these indices have been measured in the scale of 0 to 1. A value near 1 represents better provision of particular indicator in the given state while a score nearer to zero depicts dismal performance of the school.

Among these three indicators, the QTI represents the Quality of Teaching which has been constructed using a set of five variables. These variables are (as per Right To Education norms): i) pupil-teacher ratio; ii) teacher-classroom ratio; iii) number of multi-grade classes, iv) Head

Master appointed; and v) Teacher Experience. Each of these variables has been converted into an index using the following formula:

$$\text{Dimension Index of } X_i = (\text{Observed value of } X_i - \text{Minimum } X_i) / (\text{Maximum } X_i - \text{Minimum } X_i)$$

$X_i, (i=1, 2, \dots, 5)$  represents the  $i$ -th variable.

Another component of the Quality of Education Index is School Infrastructure Index (SSI), constructed using a set of 17 variables. These variables are:

1) Status of building and classroom. It is calculated on the basis of: i) whether the building is owned or rented; ii) Condition of classrooms [total no of Pucca classrooms/Total no of classrooms] iii) Separate room for Headmaster; and iv) Number of rooms other than classroom (office room/store room etc), 2) Playground, 3) boundary wall, 4) availability of drinking water facility, 5) Availability of purifying system, 6) Electricity [i) electricity available ii) functional light and fan in each room], 7) Total no. of toilets and urinals available and usable- 8) Separate toilet for girls, 9) kitchen shed, 10) Number of library books in use, 11) Teaching Learning Materials (TLM) in class IV, 12) Functional computers, 13) CAL ( Computer Aided Learning) facility, 14) game equipment in use, 15) Ramp for disabled children needed, 16) Ramp for disabled children available and 17) Trained teacher.

The third component of QEI is Learning Achievement Index /Score (LAI/LAS). The availability of any infrastructure is vital in order to impart effective and easy learning to students. Five variables have been used to check the learning effectiveness of education system: i) percentage of total marks obtained in Bengali, ii) percentage of total marks obtained in English, iii) percentage of total marks obtained in Mathematics iv) percentage of total marks obtained in

Environmental Science, v) percentage of total marks obtained in Bengali, English, Mathematics and Environmental Science based on a PRACTICAL TEST (a special test for measuring cognitive skills developed with the consultation of Educationalists and experts) conducted by the researcher on five selective questions from each subject.

Using all these five variables, LAI/LAS has been constructed as simple average of these five indices.

### **Identification of Factors explaining the Variation in LAS**

Regression analysis is used for identifying the factors (variables) that explain the variations in the learning achievement scores of the students.

These variables are qualitative and quantitative in nature. There is one quantitative variable- annual income of the family of the student concerned. All other variables are qualitative by nature. These are, for example, education level of the parents or the Head of the family, sex, caste, school management, location of the school, class attendance, school infrastructure and the like. As explained above, each of these dummy/ qualitative variables has several characteristics varying from two to four- one of them is used as benchmark or reference characteristic.

As already stated the total number of schools (primary) considered in our study is 16, out of which 8 are private schools and eight are public schools (government aided). Of these 16 schools, 8 are in the rural areas and eight in the urban area. The total number of students constituting our sample is 154. This means that the size of our sample is 154. Data on various quantitative and qualitative variables have been collected for these 154 students and 16 schools.

The dummy variables are represented by  $D_2, D_3, D_4$  etc. and the quantitative variable, i.e., the annual income of the family, by  $X_2$ .

The regression model is of the form as given below:

$$y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \dots + \beta_k D_{ki} + \gamma_2 X_{2i} + u_i \quad (i=1,2,\dots,n)$$

Where  $D$ 's are dummy variables and  $X$  is the quantitative variable,  $U$  is the disturbance term,  $\beta_2, \dots, \beta_k$  are the slope differentials,  $\beta_k$ , for example, measures the difference in the amount of influence of the  $D_k$  dummy variable over the base or the reference variable or characteristic,  $\gamma_2$  measures the effect of  $X_2$  on  $Y$  variable.

Similarly, the model for identifying the determinants of school enrolment of the students is of the same form, which includes both the quantitative and qualitative variables.

Some of the explanatory variables are the same as several variables in the learning achievement score and others different. The following explanatory variables are assumed to have influence on the students' enrolment in different schools. These are: 1) Academic qualifications of the parents of the students ( $X_2$ ), 2) Average annual income of the household ( $X_3$ ), 3) School fees ( $X_4$ ), 4) Learning Achievement Score ( $X_5$ ), 5) Quality of Teaching Index ( $X_6$ ) 6) School Infrastructure Index ( $X_7$ ), 7) female-teacher ratio ( $X_8$ ), 8) location of the school ( $D_1$ ), 9) School management or organization ( $D_2$ ), 10) Availability of UP (Upper Primary) and or Secondary or Higher secondary schools within 1 to 2 k.m. of the primary school (s) ( $D_3$ ) where the students seek admission school, and 11) Headmaster as a leader ( $X_9$ ).

However, though all these variables appear to have good influence on the student enrolment in schools, the number is as many as eleven where as the number of observations is just 16 i.e., 16

schools for which the observations on these variables are collected and for which the used for estimating the parameters associated with these eleven variables. The degree of freedom is  $n-k-1=16-12=4$ . It is quite likely that given this value of dfs, most or all parameters will be not at all significant even at a very high risk level i.e., at a high level of significance, say 10 per cent. Further, the variables are of such nature that many of them may be highly correlated and the correlation coefficient may be statistically significant at very low probability levels.

The problem of non-significance of the parameters mentioning the model useless may be solved to a large extent by increasing the number of freedom. One way of doing this is by increasing the size of the sample in our study. The other way is to eliminate some of the explanatory variables from the model. We consider the second way since the first method is out of consideration at this stage of work. This method is used by following the steps mentioned below.

Step 1: we regress the dependent variable – school enrolment index of the students ( $y$ ) on all the independent variables mentioned above.

Step 2: We check the estimates of the coefficients of the variables, their  $t$ -values and  $p$ -values along with the nature of relationship (+ or –) between the dependent variable and the independent variables in the multivariate relationships. It may so happen that some of the parameters are statistically insignificant.

Step 3: If it happens so, we check the pair-wise simple correlation among the independent variables and correlation between dependent variable and each one of the independent variables.

Step 4: To start with we select those independent variables which have significant correlation with the dependent variable. Then check the degree of association between the independent variables and their levels of significance.

If there is significant association/correlation between any pair of independent variables, then we eliminate one of these two and retain the other. We retain that variable which has higher association /correlation with the dependent variables in relation to that of the other independent variable. For example, we have a 4 variable multiple regression model:

$$y = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$$

We run this model by applying OLS method of estimation of the parameters  $\beta$ 's, and we estimate  $\beta$ 's. Let the estimates of the  $\beta$ 's and then t-values and p-values be

$$\hat{\beta}_1 = 16.50$$

$$\hat{\beta}_2 = 13.93 \quad t = 1.45 \quad p = 0.151$$

$$\hat{\beta}_3 = 8.09 \quad t = 1.45 \quad p = 0.234$$

$$\hat{\beta}_4 = 10.23 \quad t = 1.23 \quad p = 0.126$$

and,  $r_{x_2x_3} = 0.67, \quad r_{x_2x_4} = 0.61 \quad r_{x_3x_4} = 0.50$

$$r_{yx_2} = 0.75 \quad r_{yx_3} = 0.43 \quad r_{yx_4} = 0.64$$

$x_2$  – included;  $x_3$  – excluded;  $x_4$  – included;  $N = 7$ ;  $df = 7-4 = 3$

The above results for model  $\hat{y} = 16.50 + 13.93 X_2 + 8.09 X_3 + 10.23 X_4$  show that none of the variables have any significant influence on the dependent variable Y. These results under the

model are useless. The uselessness of the model may be due to very small simple size ( $N = 7$ ) and small degree of freedom  $\{N-(K+1) = (7-4) = 3\}$  where  $K$  is the number of explanatory variables.  $K+1$  = number of explanatory variables plus constant. To increase the number of degree of freedom we have to eliminate as many explanatory variables without significantly losing the explanatory power of the model.

The above results help us identify the explanatory variables which can be retained and which should be eliminated.  $X_2$  and  $X_4$  be retained and  $X_3$  eliminated, since

- 1)  $r_{yx2}=0.75 > r_{yx3}=0.43$  and  $r_{yx4}=.64$  and  $r_{x2x3}=.67$  so  $x_2$  is retained and  $x_3$  is eliminated.
- 2)  $r_{yx4}=.64 > r_{yx3}=.43$  and  $r_{x3x4}=.54$ , so  $x_4$  is retained and  $x_3$  is eliminated.

So following this procedure we are left with model  $\gamma = \beta_1 + \beta_2 x_2 + \beta_4 x_4 + x$ ,  $N=7$ ,  $df=7-3=4$

$df$  increases by 1. If the model is a 16 or 17 variable model, elimination of the relatively insignificant variables will increase degrees of freedom significantly. This procedure leads to the presence of quite a few explanatory variables that will have statistically significant effect on the dependent variable. This is shown in our regression analysis of the school students' enrolment on the explanatory variables mentioned above as well as in the students' learning achievement score (LAS) models developed to identify the factors or variables that significantly influence these two regressands – school students' enrolment and students' learning achievement score.