Ministry of Education
Government of India

सत्यमेव जयते

विद्यया ऽमृतमश्नुते

एन सी ई आर टी
NCERT

FLS
2022

Sampling Design and Sample Selection for Foundational Learning Study (FLS)

Supported by
unicef
for every child

Sampling
Design and Sample Selection for Foundational Learning Study (FLS)

Preface

Education is the cornerstone of progress and prosperity for any nation. It is through education that we empower individuals to unlock their full potential and contribute meaningfully to society. Recognizing the transformative power of education, the Government of India launched the National Initiative for Proficiency in Reading with Understanding and Numeracy (NIPUN) Bharat in July 2021. This visionary initiative, aimed at ensuring that all children attain foundational skills by the academic year 2026-2027, underscores our commitment to providing quality education for every child, regardless of their background or circumstances.

At the heart of the NIPUN Bharat initiative lies the Foundational Learning Study (FLS), a comprehensive assessment conducted by National Council of Educational Research and Training (NCERT) across India. The Foundational Learning Study, undertaken in March 2022, represents a crucial step towards understanding and addressing the learning needs of Grade 3 students across India.

The primary aim of the Foundational Learning Study was to draw a first-hand understanding of students' learning levels at the foundational stage, with a specific focus on two key goals: effective communication and informed learning. Through a series of tasks designed to assess foundational literacy and numeracy, the study sought to identify areas of strength and areas for improvement, thereby informing targeted interventions and policy reforms.

One of the defining features of the Foundational Learning Study was its inclusive approach, with assessments conducted in 20 languages to ensure representation and accessibility for all learners. This commitment to linguistic diversity reflects our belief in the importance of providing education in the language of the learner, thus fostering greater inclusivity and cultural relevance.

Following the administration of the study, a comprehensive benchmarking exercise was undertaken to contextualize students' performance levels and set language-specific benchmarks for oral language fluency and reading comprehension. These benchmarks, informed by empirical data and best practices, serve as essential reference points for educators, policymakers, and researchers as they work towards improving educational outcomes.

The dissemination of the Foundational Learning Study findings at both the National and State levels mark the beginning of a new phase in our educational journey. This report offers detailed insights into the nuances of foundational learning, providing a roadmap for targeted interventions and policy reforms.

This prodigious task could not have been achieved without the active partnership and association from all the SCERTs, SIEs and the SPDs office of the different States and UTs. The completion of this astounding work also entailed a dynamic technical collaboration with UNICEF, India. I would like to express my gratefulness to Mr. Terry Durnnian, Chief, Education, Mr. Ramchandra Rao Begur, Education Specialist, Mr. Ganesh Kumar Nigam, Education Specialist and Ms. Sunisha Ahuja, Education Specialist.

As we embark on the next leg of our journey towards educational equity and excellence, let us remain steadfast in our commitment to providing every child with the opportunity to realize their full potential. Together, we can build a future where education is not just a privilege but a fundamental right for all

Prof. Indrani Bhaduri

Head ESD, NCERT

Sampling Design and sample selection for Foundational Learning Study (FLS)
March 2022

1. Introduction

Foundational Learning Study(FLS) is a nationally representative large-scale survey of students' of grade 3 on reading proficiency in language in India administered first time to set reading benchmark. In the 5th meeting of the National Steering Committee (NSC) for the FLS study held on 10th January, 2022 decided that, FLS Survey will be administered to Grade 3 in case of States where the academic session for class 3 is ongoing and for states where the academic session for class 3 has ended children entering class 4, students who just entered in Class 4 will be taken as part of the sample.

Sampling design and procedures play a crucial role in ensuring that the results obtained on a sample can be reliably applied to the entire population. Sampling design for the FLS 2022 intends to support the predefined and agreed ***objectives*** of the national assessment. FLS 2022 aims to provide information of what India's students of Grade 3 know about language and how proficiently they read in the language which is Medium of Instruction in the respective schools. The assessment will be for Grade 3 (or new admission to Grade 4 if next academic session started on the date of exam), which can set benchmark at ***National***, ***State***, ***Medium of Instruction(MoI)*** and ***School Management*** levels, as schematically depicted in Figure 1. Since, FLS 2022 is being the first of its kind Nation Wide Survey, Kendriya Vidyalaya (KVs) is taken as separate management unit separating from Government.

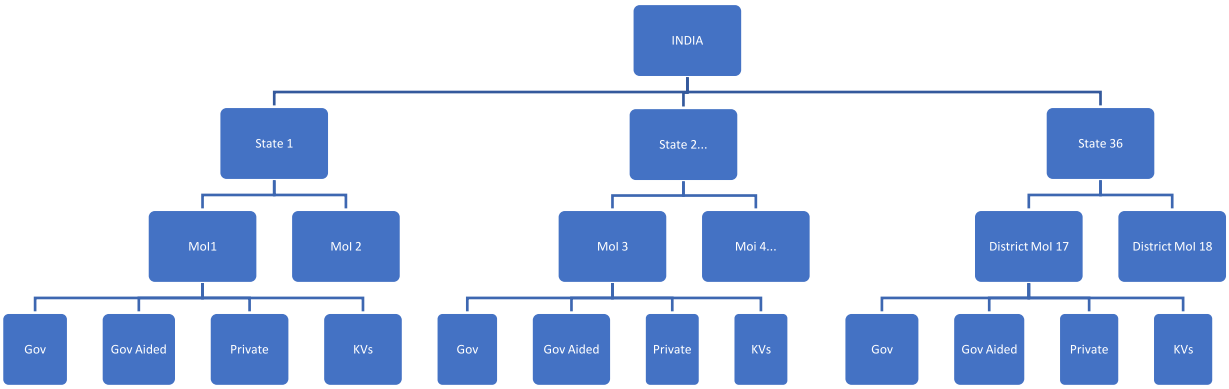

Figure 1 Hierarchical structure of the ORF 2020-22 sample strata

Sampling plan for FLS 2022 is designed to enable reporting of Grade 3 (or new admission to Grade 4 if next academic session started on the date of exam); student's reading performance at each of the hierarchically organized administrative levels, as presented in Figure 1, in modern Indian languages. Thus, FLS 2022 will inform about students reading proficiency in grades 3 at overall national level, how do they perform in each of the States, and what is their achievement within each of the Indian language and lastly, within each language, student proficiency will be reported for each of the school administration types (government, government aided, private and KVs). A large scale sample survey involves five-step process for selecting a sample and administering a survey viz., defining the population, specifying the sampling procedure, determining the sample size, selection of the sample and finally administering the survey.

To achieve the objectives, there are several questions that the sampling design will address:

1) What is the Sampling frame?
2) What is the minimum sample size needed to obtain reliable information of student performance in the groups at the lowest stratum level – school types within districts?
3) What are the resultant sample sizes at district, state, and national level?
4) What methods will be used for selection of schools within each school type and for selection of students within schools?

In the next sections we present a rationale of the sampling design and factors relevant for answering these questions.

2. Frame to be used for Survey:

The NSC set up for finalizing procedure for ORF Study 2022 in their 3rd meeting held on 10th November, 2021 has deliberated on various options of sampling frame and availability of latest data such UDISE+ 2020-21 etc and noted that National Achievement Survey (NAS) 2021 has been conducted successfully in November 2021. NAS 2021 used UDISE+ 2019-20 as frame. Accordingly, it was decided that, a subsample to be selected from final sample of NAS 2021 for the purpose of FLS study 2022, and extended to NAS 2021 frame, in case of shortage of required sample number from final NAS sample. It was also noted that, though the schools will remain the same in FLS Study 2022, but the students would not be the same providing enough randomization. Therefore, the method, sampling design and selection of sample followed for the NAS 2021 sampling has been adopted in FLS also. However, main steps followed in sampling design and sampling frame of FLS is elaborated in the subsequent paras.

3. Logic of Sampling

In a typical large-scale survey, data are obtained on a sample of students drawn from the population. Considering that the results based on sample data can be generalized to the population only under certain conditions, it is important to identify required characteristics of the sample that enable attribution of the statistical information obtained on a sample of students from a certain group to all the students in that group.

The relationship between sample and population is based on a statistical probability, which means that descriptions and inferences about the characteristics of the student population based on a sample data are never fully accurate; they are made only to some degree of certainty. It is just a matter of statistical conventions to determine what is considered as a sufficient ***degree of certainty*** at which we should accept that the characteristic observed on a sample can be applied to the population. This convention is commonly called “statistical significance”.

Thus, the focal points of sampling are the factors associated with this degree of certainty and how do we ensure that a sufficient level of certainty will be reached?

The answers are derived from the theorems of normal distribution and the concept of “standard error” that is associated with every statistical indicator derived from sample data. The size of standard error is proportionate to the variability of measured phenomenon and inversely related to the sample size. Thus, with increase of sample size the standard error decreases, which further leads to the main sampling question – ***how large should be sample size*** so that standard error becomes sufficiently small to enable the ***desired level of certainty*** in making conclusions about student performance in the population? The answer to this question is dependent on two categories of factors:

a) Statistical conventions and deliberations that we accept to abide.
b) Objectives of the assessment, i.e., what information do we intend to derive from data.

4. Statistical conventions and deliberations

The main statistical conventions and deliberations that need to be defined and agreed upon when deciding about sample size include:

1) ***Statistical significance***: Defines the probability at which we decide that results observed on a sample can be generalized to the population. This probability can be interpreted as the level of risk associated with accepting the hypothesis that the statistical information derived from a sample can be generalized to the population. This probability is also called Type I error and it is conveniently set to 0.05 (symbolized as alpha), which means that we accept a risk up to 5% that we are wrong when rejecting a hypothesis of chance, and generalizing the effect observed on a sample to the entire population.
2) ***Statistical power***: Defines the probability at which we can decide that "no-effect" observed on sample can be generalized to the population. The complement of power (1 – power)is also called Type II error and it is typically set to 0.20 (symbolized as beta), which means we take a risk of up to 20% that we are wrong when stating there is "no-effect" and generalizing this finding to the population.
3) ***Minimum detectable difference***: Size of the difference that we can detect based on data collected from our sample. This difference may refer to either the difference between a sample statistic and a population parameter (e.g., average student performance), or the difference between two or more groups of students (e.g., boys vs. girls). Minimum detectable difference (MDD) can be standardized, i.e., expressed as a fraction of standard deviation, in which case the value is called ***minimum detectable effect size*** (MDES=MDD/s). Defining a desired MDES affects the sample size estimation because larger samples are needed for detection of smaller differences. The MDES for the FLS 2022 sampling is set to be 0.28, which means that the sample size needs to be sufficiently large to detect a difference of the size equal to 28% of standard deviation.
4) ***Cluster sampling method***: In educational research sampling units (students) are grouped in clusters (schools). Since the utilization of Simple Random Sampling (SRS) would be associated with logistic challenges, it is common practice to use a cluster sampling (CS) method. In CS we first select a sample of clusters (schools) and then a sample of students within each school. Since the schools share common contextual factors, there is a certain degree of similarity or uniformity of student performance within schools. Consequently, increasing the number of students within schools does not bring much of new information, which requires a larger sample size as compared to SRS method. Thus, depending on the level of uniformity of students within schools, the estimation of sample size should be increased by an inflation factor called ***design effect (DEFF)***. This adjustment is based on the level of similarity between the members within clusters, technically called ***intra class correlation (ICC)***, which is calculated from existing data collected by similar instruments in similar conditions.
5) In addition to statistical conventions and deliberations, sample size is also informed by objectives and nature of the assessment information that is intended to be derived from data. FLS 2022 aims to provide information of what India's students of Grade 3 know about language at the national level, state level and management level and how proficiently they read in the language which is medium of instruction in the respective schools. When research questions require deep disaggregation of the sample down to the management

level, then several key questions need to be considered as they will substantially affect the sample size at all levels.

A. What is a ***minimum size of the group that can be meaningfully reported*** with a reasonable level of precision? Examples: estimating average district performance score, estimating how this score relates to some standard (performance level cut score, or national average) or any other fixed value. This requires sample size estimation method for one group vs. population (or fixed value).
B. What are ***minimum sizes of the groups that can be compared*** at a reasonable level of power? Examples: comparisons between girls and boys, urban and rural students, students from different social groups, or students from different school management). This requires sample size estimation method for two independent groups.

5. Method A: one group vs. population comparison

The estimation of sample size based on method A (one group vs. population or fixed value) relies on evaluation of standard error of the mean and associated confidence intervals based on central limit theorem and sampling distribution of the sample mean. In the case of one group, there is one standard error as the other value being compared is fixed (for example cut score) or associated with negligible error (group population mean, national mean).

A common formula for estimation of sample size for SRS in the case of one group *Vs* population is derived from the expression for Confidence Interval:

$n = 4\left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{s^2}{(2\ MDESs)^2}$ which simplifies into $n = \left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{1}{MDES^2}$

6. Method B: two groups comparison

Commonly used formula takes the following inputs: Type I and II error rates, variance, minimum detectable difference, and the relative sizes of the compared groups (most commonly they are treated of being equal). Total sample size for both groups is estimated by the following expression:

$$N = \left(\frac{1}{q_1} + \frac{1}{q_2}\right)\left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2 \frac{1}{MDES^2}$$

Where q_1 and q_2 are proportions of cases in group 1 and group 2 (typically 0.5 for both if the groups are of the same size). Then, a sample size for each group equals:

$n_1 = q_1 N$ and $n_2 = q_2 N$

It is important to consider this formula for ORF 2022 in situations where two compared groups are of unequal size, for example by school management, government and government aided school groups may be of substantially different size if proportionate sampling is used. The alternative is to use non-proportionate approach (which in the stage of data analysis needs to be compensated by sampling weights) and to set the groups to equal size.

7. Adjustments of sample size

7.1 Based on size of finite population

Further adjustments of sample size can be done in case the size of finite population is known. This adjustment does not have effect for larger populations, but for smaller populations the estimated sample size will be progressively decreased.

The following formula can be used for adjustment based on the known finite population size:

$n_{adj} = N*n / (N + n - 1)$

where n_{adj} is adjusted sample size, n is original sample size, and N is size off infinite population.

7.2 Based on design effect

The rationale for adjustment by DEFF is described earlier in this document and is normally used in practice whenever cluster sampling method was applied. DEFF is typically calculated from similar data administered in similar conditions. Researchers are cautioned to be careful in making decisions about the size of DEFF as it is not certain to which degree are old data alike the new data to be collected. Analysis of NAS 2017 data to obtain the ICCs at district levels showed a large variance among districts in estimating ICCs and subsequently DEFF. The range of DEFF was between 2 at lower end and 20 at high end, with most of values being in the range between 5 and 10. Overall, these are large DEFF factors and their application sets very demanding sample sizes. For the purpose of FLS 2022 sample size estimation, the ICCs of 0.6 assumed. In FLS survey, it is planned 10 students will be assessed per class/ section that means ***the cluster size is 10*** and assumed ρ(ICC) value is 0.6 (this is first FLS survey). So, the Design effect for this survey comes out to be 6.4. This means that the estimated sample size for SRS will be multiplied by factor 6.4 to obtain sample size for CS.

7.3 Based on non-response rate

The adjustment of sample size based on the non-response rate compensates the anticipated absenteeism of students based on previous experience with similar assessments. For FLS 2022 it is determined that the sample size will be increased 15 % to compensate non-response rates.

8. Determination of sample size for FLS 2022

Step 1: Finalize coverage of Schools:

1.1. Frame (all classes): UDISE+ 2019-20 will be used which is same as NAS 2021 sample frame. A sub sample is drawn from NAS 2021 selected sample.

1.2 Target population: Students of Class 3 of all recognised schools of Government, Government aided, private and Kendriya Vidyalaya (KVs) **(except Ministry of Labor, Jawahar Navodaya Vidyalaya, Sainik School, Railway School & Central Tibetan School)**

1.3 Mediums of Instructions (MoI) under coverage for each class: 20

Assamese, Bengali, English, Gujarati, Hindi, Kannada, Malayalam, Manipuri, Marathi, Mizo, Odiya, Punjabi, Tamil, Telugu, Urdu, Bodo, Garo, Khasi, Konkani and Nepali .

1.4 Exclusion of small schools for administrative convenience: If enrolment in a class is less than 6, school will be excluded for sample selection of that class.

Computation using UDISE+ 2019-20	Grade-3	
Total Class 3 Enrolment	2,43,04,294	A
Out of Frame (Management 8 and 98)	6,22,688	B
Desired national target population	2,36,81,606	C=(A-B)
Exclusions :		D
Small schools (enrolment less than 6 in grade 3)	7,73,032	D
Ministry of Labor, Jawahar Navodaya Vidyalaya, Sainik School, Railway School & Central Tibetan School	8273	**E**
Enrolment of Visually impaired(CWSN)	34079	**F**
Enrolment in mediums of instructions not covered	3558	**G**
Total Enrolment, who are excluded	788421	H=(D+E+F+G)

% of excluded enrolment	3.34	
Defined national target population	2,28,62,664	I=(C-H)

1.5 Sampling frame structure: Sampling frame will contain the following columns:

- State Code
- State Name
- District Code
- District Name
- Block Code
- Block Name
- School Code
- School Name
- Management Code
- Management (Govt./Aided/Private Unaided/ KVs)
- Category Code
- Location (Rural/Urban)
- Enrolment Boys
- Enrolment Girls
- Enrolment Total
- Total Enrolment in that particular Medium of Instruction (MoI)
- Whether selected grade have more than one medium of instruction (Yes/No)
- Whether selected grade have more than one section (Yes/No)
- Whether selected grade have any children with special needs (Yes/No)

Step 2: Make separate frame of each broad type of school management or Stratification as under: (school management codes as per UDISE+ Data Capture Format)

Frame 1: State Government schools (coverage of NAS 2021)	School management 01, 02, 03, 06, 90 (01-Department of Education, 02 – Tribal Welfare Department, 03 – Local Body, 06 – other Government managed, 90 – social welfare department)
Frame 2: Government Aided schools (coverage of NAS 2021)	School management 04 (04 – Government Aided)
Frame 3: Private Unaided recognised schools (coverage of NAS 2021)	School management (05- Private Unaided Recognised, 97 – recognised madrasa)
Frame 4: Kendrya Vidhyala schools (coverage of NAS 2021)	School management 92

The objective is to provide separate estimate of MoI at these school Managements.

Step 3: Estimating target sample size per MoI per Management

Inputs: assumed Confidence level 95%, assumed intraclass correlation (ρ) value is 0.6(same as NAS 2021), cluster size of 10, Minimum Detectable Effect Size = 0.28, the calculated Design Effect =1 + (10-1) X 0.6=6.4.

Output: Estimated sample size per MoI = 314.

Adjusted sample size per MoI test with non-response adjustment rate of 15 % is 361 (approximated to 365). Therefore, the minimum sample size for any subject was decided as 365.

Step 4: Adjusting of sample size determined in step 3 (apply for less than 10,000 students):

$n_{adj} = N*n / (N + n - 1)$

Examples:

i. **if enrolment of class 3 of Government schools in a State is 5000, target sample size for MoI of class 3 will be = 5000 * 314 / (5000 + 314 – 1) = 296**

ii. **if enrolment of class 3 of Government schools in a State is 1000, target sample size for MoI of class 3 will be = 1000 * 314 / (1000 + 314 – 1) = 239**

Step 5: Deciding target sample number of students per Management for frame 1 to 3

Number of subjects to be assessed at given grade divided by number of subject tests taken by each student is used to get multiplier based on assessment design. Since, each student is tested only on one **Mediums of Instructions (MoI)** , each selected student will undergo only one test, therefore, **Multiplier based on assessment design is 1 and Sample size per stratum after Step 4 = 365 * grade-wise multiplier = 365*1=365**

Step 6: Selection for Frame 1/2/3: State

Sample size requirements:

A. Check the minimum student requirement for each MoI for each State. Thus, if number of students in the State is less than or equal to 365 as per frame, all schools in the frame will be selected for survey for that MoI.

B. Maximum number of students to appear in test from a school is 10(. Therefore, compute maximum effective sample students from all schools, i.e., Minimum (grade enrolment, 10). If sum of maximum effective sample student is less than or equal to 365, select all schools in the frame.

For the state "X", MoI where enrolment is higher than those described in steps 6-A and 6-B above, sampling will be necessary.

Size measure: Total enrolment in the target class will be considered as size measure (i.e., when we select schools for Class 3, enrolment for class 3 as per UDISE+ 2019-20 will be considered as size measure).

Step 6-C: (for States not covered in Step 6-A or 6-B)

i. Suppose E is total enrolment and S is total schools. Then, compute round (E/S, 0) as average enrolment of school (AES).
   a. If AES <=10, find target number of sample schools TS_1= (365/AES). To select TS_1 schools, take sampling interval (SI) = round(E/ TS_1 , 0)
   b. If AES >10, find target number of sample schools TS_1= integer (365/10) +1 = 37. To select TS_1 schools, take sampling interval (SI) = round(E/ 37 , 0)

ii. Select all schools with enrolment > SI with probability 1. This is done to ensure that such large schools do not get re-selected in the systematic sample. Then, re-compute SI with reduced number of schools and enrolments.

iii. Repeat Step 6-C-i and 6-C-ii till enrolment of all remaining schools are less than or equal to SI (or modified SI as per re-computation done in step 6-C-ii).

iv. Suppose S_1 schools have been selected in Steps 6-C-i to 6-C-iii.
   a. Enrolment of these S_1 schools is E1.
   b. Suppose some of these schools have more than 10 students in a class, while remaining schools have less than 10 students. Assume, from these S_1 schools, we get a target sample of n_1 students.
   c. Therefore, we have to further select (365 - n_1) target students from (S - S_1) schools.

v. Now, AES_1 = (E - E_1) / (S - S_1), the average class size of (S - S_1) schools. Modified target number of sample schools will be TS_2 = (365 - n_1) / AES_1 . Compute SI_1 = round((E - E_1) / TS_2 , 0)

vi. Compute X_1 = (365 - n_1) / SI_1.

vii. From the remaining schools of the district, select X_1 schools by PPS circular systematic sampling with SI_1 as sampling interval.

viii. For circular systematic sampling, the schools in the frame will be arranged by enrolment in the class (descending), medium of instruction (if a school has enrolment in more than one medium of instruction, the medium in which grade enrolment is maximum will be considered) and area (rural/urban). After that, prepare a column with cumulative frequency (using enrolment in the specific class/grade). To drawn the samples using circular systematic sampling, draw a random number between 1 to (E - E_1). Let this be RN_1. Select the school against which RN_1 falls as per the cumulative frequency table (example given below). Then, compute (RN_1 + SI_1), (RN_1 + 2*SI_1), …, (RN_1 + (X_1 - 1) * SI_1), to earmark the remaining sample schools. During this phase, if (RN_1 + i * SI_1) becomes higher than (E - E_1), compute ((RN_1 + i * SI_1) - (E - E_1)) to get the number for school selection.

ix. After selection through Steps 6-C-v to 6-C-viii, if it is found that the effective number of sample students is less than (365 - n_1), then Steps 6-C-v to 6-C-viii will be done afresh, starting with increasing the target number of sample schools from TS_2 to TS_2 + 1.

Step 7: Selection for Frame 4 (Kendriya Vidyalaya (KVs))

Note that total enrolment in KVs of class 3, as per UDISE+ 2019-20 is 1.158 lakhs, which is only 0.51% of enrolments in schools under FLS coverage. Total KV schools are **1252** at all-India level with representation at all states. In 2 states there are more than 100 schools each and another 19 states have more than 34 KVs each. In 15 States, total KV schools are less than 25. Most of the KVs have the MoI as English in few cases Hindi . Therefore, KVs are selected for 2 MoIs only.

i. If number of KVS in a State is 25 or less, for selected MoI, then these schools are not selected as the required minimum sample requirement is not met

ii. If number of KVS in the State for selected MoI is more than 25, schools will be selected from the State by PPSWR, with total enrolment as the size measure.

Step 8: Selection of Section and Students in a selected school (All the frames same as NAS 2021 method)

i. In a selected school, if the enrolment in target class is less than 10, all students will be selected for the test.

ii. If enrolment in target class is more than 10 and students are in a single Section and number of students present on the date of assessment is more than 10, 10 students will be selected at random using circular systematic sampling based on class attendance register. The total number of students in the class should be divided by 10 to get the interval (m, rounded off to the nearest lower integer). The first number can be decided based on draw of lot method. Suppose the class has 45 students. Then, numbers 1 to 45 can be written on small pieces of paper, folded alike, and mixed well. Then one piece will be drawn randomly. Suppose the number drawn is 35. This will be the roll number of the first random student selected for the test. Then, every m^{th} student should be selected starting from roll number 35 in the attendance register, till selection of 10 students is complete. If some selected roll numbers are absent on the day of test, then the selection process will continue till a total of 10 students are selected for the test.

iii. If students are in multiple Sections, the following step-wise procedure will be followed:

a. One section will first be selected at random. This can be done through draw method. Names/ numbers of all sections of the class can be written on small pieces of paper, folded alike, and mixed well. Then one piece will be drawn randomly and that would be the Section of the class where students will be assessed. The Teacher Questionnaire will be canvassed for this selected section only.

b. Up to 10 students from this selected Section will be selected for the test by circular systematic sampling as described above in step 8_ii.

c. Suppose the number of students on the date of examination in the Section selected at Step 8-iii-a is n_2 (less than 10), one more Section will be selected by SRSWOR.

d. $10 - n_2$ students will be selected by circular systematic sampling from the Section selected at Step 8-iii-c.

e. Step 8-iii-c and 8-iii-d will be continued till a total of 10 students are selected from this class in the selected school. This is to be done to minimise loss in target number of sample students.

f. This get repeated for selection of more than one section if sample drawn from the school is more than 10 to meet the desired size.

Iv Alternatively, for selection of section method followed in NAS 2017 by NCERT may also be adopted i.e if there are more than one section in class 3 of the selected school and only one section is to be selected for FLS, then select the PIN code of the school to find the random number and add all the digits of the pin code to get the random number. Then write down all section in a line then start from 1 then move to next section in a cyclical way till the identified random is achieved. Stop once the identified random is reached and mark it as selected section for study. Repeat same if more than one section is to be selected.

V. Detailed list of selected sample of Students for FLS based on the above procedure is at Annexure I.

10. Additional Sample: To meet 1200 in each MoI at state level:

In the National Steering Committee meeting held on 17.2.2022, UN representative / AIR has informed that, four different reading passages with different lengths and difficulty are included

in FLS by NCERT. These passages of different difficulty levels need to be brought on the same measurement scale. This is required to create a conversion table between these 4 passages i.e., wherein, 5 score points on an easy passage is equivalent to 1 score point on a difficulty passage. Therefore, it is necessary to carry out chained linking. This will bring 4 passages on the on the same measurement scale. To chain optimally, variation in data is needed and to achieve this variation, a minimum of 300 data points on 1 booklet is needed. Overall, across 4 booklets, 1200 data points will be needed.

To meet the requirement of 1200 students per state, additional sample selected in the MoI's falls short of 1200 in a State. The selection process is same as the one elaborated the preceding sections. After the selection of samples State/ MoI wise, it is noted that eight MoIs where found to be short of requirement of 1200 students. It is further noted that, total schools / students available for additional selection is very limited. As a result, maximum number of students per schools has gone up from 10 in the original selection to all available students in many cases and even after this required 1200 is not met in case of Manipur and Mizo at State level. The details of additional students to be drawn from original selected schools are as under

State	Medium of Instruction	State total in MoI in the original selection	Additional selection required	Selected additionally	Enrolments Shortfall to reach 1200 at state level
Assam	Bodo	742	458	465	Nil
Meghalaya	Khasi	1118	82	85	Nil
Goa	Konkani	369	831	835	Nil
Kerala	Malayalam	1122	78	80	Nil
Manipur	Manipur	367	833	156	677
Mizoram	Mizo	746	454	353	101
West Bengal	Nepali	742	458	460	Nil
Punjab	Punjabi	1116	84	85	Nil

Detailed list of selected sample to meet requirement of 1200 students in a State in a given MoI (Total 8 MoI's in 8 States). Therefore some of the selected schools in this list might be there in the original list also as there is a shortage of total enrolment to select the schools independently. Consequently, more than 10 students may be there in some schools under this category and the detailed list is at Annexure III.

11. Additional Sample: For special 10 districts of select 10 states:

To study the learning level in the selected district of 10 viz., Changlang, Dhamtari, Diu, Alirajpur, Yavatmal, Sambalpur, Muktsar, Tiruchirappalli, Adilabad and Shamli, additional samples were selected over and above the initial selection. Therefore some of the selected schools in this list might be there in the original list also as there is a shortage of total enrolment to select the schools independently. Consequently, more than 10 students may be there in some schools under this category and the detailed list is at **Annexure II**

References:

1. World bank, Sample size and power calculations https://dimewiki.worldbank.org/wiki/Sample_Size_and_Power_Calculations
2. Rosner B. (2011) *Fundamentals of Biostatistics*. 7th ed. Boston, MA: Brooks/Cole.