CENTRAL SQUARE FOUNDATION

Design and Implementation of State-wide FLN Assessments

July 2025

Table of Contents

Acknowledgements

This playbook on the Design and Implementation of State-Wide Foundational Literacy and Numeracy (FLN) Assessments has been developed through the collaborative efforts of the Research, Monitoring, Evaluation, Assessment, and Learning (RMEAL) team. We acknowledge the technical expertise and contributions of Shruti Bhadbhade, Dr. Geetika Bhandari, Susmit Biswas, Alankita Upadhyay, Aniket Marne and Debapriya Chanda in shaping the content, tools and implementation approach presented in this document. The work was carried out under the guidance of Akashi Kaul, Director, Research, Monitoring, Evaluation, Assessment and Learning (RMEAL) team, whose strategic direction and insights greatly strengthened the playbook.

We are deeply grateful to our CEO, Shaveta Kukreja, for her leadership and unwavering support throughout the development of this playbook. Her vision for strengthening foundational learning has been a constant source of inspiration.

Finally, we thank all members of the development team, implementation teams, and reviewers whose inputs were invaluable in shaping the structure and substance of this resource.

Foreword

Assessments play a pivotal role in the success of the Foundational Literacy and Numeracy (FLN) mission in India. They are critical tools to gauge the effectiveness of educational interventions and measure the progress of students' literacy and numeracy skills.

Through reliable and standardized large-scale assessments, policymakers and stakeholders can make data-driven decisions, directing resources where they are most needed and ensuring the FLN mission's objectives are on track. Large-scale assessments also act as a mechanism for accountability, encouraging transparency and continuous improvement in the education system. Moreover, a well-designed and administered large-scale assessment can provide regular, system-level information on student learning to gauge overall levels of achievement and the performance of specific sub-groups that may be at higher risk of falling behind. If data from such an assessment is analyzed and released in a timely and useful manner, the government can determine the effectiveness of its policies and alter them accordingly.

Creating an effective large-scale FLN assessment requires meticulous planning, collaboration, and a profound understanding of standardized assessment design principles. This playbook, crafted by experts, provides a comprehensive guide, step-by-step guidance, and best practices from successful state-wide case studies. It paves the way for states or their users to design and implement these assessments efficiently, while also addressing operational challenges and offering key solutions for the challenges. The playbook overall demonstrates the process of planning, allocating resources, establishing timelines, setting budgets, and executing each step within the assessment cycle. As technology continues to transform education, the playbook explores innovative ways to integrate digital tools, ensuring a seamless and secure administration of assessments while maintaining data reliability and privacy. The playbook serves as crucial pre-reading material that will thoroughly orient the users before they initiate assessment preparations. It offers valuable insights and guidance for the assessment planning process, empowering users with standardized procedures and protocols. As a public good, the playbook stands ready to support and facilitate the planning and implementation of FLN assessments, ensuring a transformative impact on early education.

List of Figures

List of Tables

1 OVERVIEW

1

Overview

1.1 Introduction to FLN Mission in India

Foundational learning skills are the bedrock of education, as they are the fundamental building blocks upon which a child's future learning is built. These skills encompass the ability to read with comprehension and perform basic mathematical operations. Failing to acquire these essential foundational skills leaves a child ill-prepared to tackle the more complex curriculum beyond Grade 3. In adulthood, these skills are important for individuals to actively participate in society, pursue further education and contribute effectively to the economy.

Acquisition of foundational skills ensures quality education, which has been adopted globally as Goal 4 of the SDGs by the UNDP in 2015. Research states that "it is learning rather than years of schooling that contributes to a country's economic growth: A 10 percent increase in the share of students reaching basic literacy translates into an annual growth rate that is 0.3 percentage points higher than it would otherwise be for that country" *(Hanushek, E. A., & Woessmann, L. 2012)*. Findings like these underscore the profound impact of education on society's development and highlight the importance of fostering effective learning outcomes in the early grades for individual and collective prosperity.

In recognition of the critical role played by foundational skills in national development, the Government of India launched a national mission under the **Atmanirbhar Bharat** campaign whose primary objective is to ensure that every child in the country acquires foundational literacy and numeracy skills by the end of Grade 3, by 2026-2027. The Ministry of Education (MoE) has further launched an important initiative called **NIPUN Bharat** in July 2021, which outlines clear priorities and actionable strategies for all States/UTs to accomplish the ambitious goal of ensuring proficiency in foundational literacy and numeracy for every child by the end of Grade 3.

While 2027 marks the immediate target year, the focus on foundational literacy and numeracy will continue to remain a national priority, ensuring sustained support and progress in early learning outcomes for all children.

"The highest priority of the education system will be to achieve universal foundational literacy and numeracy in primary school by 2026-2027. The rest of this Policy will become relevant for our students only if this most basic learning requirement (i.e., reading, writing, and arithmetic at the foundational level) is first achieved." — *Chapter 2: Foundational Literacy and Numeracy - An Urgent & Necessary Prerequisite to Learning, Page 8, NEP, 2020*

The implementation of the FLN mission serves as a crucial stepping stone in the journey towards enabling every child to overcome learning challenges and develop essential skills for future academic success. The success of the FLN Mission, therefore, relies significantly on the active involvement of States and UTs. Several States have taken steps to implement the NIPUN Bharat Mission. Some of these missions have been highlighted in the image below.

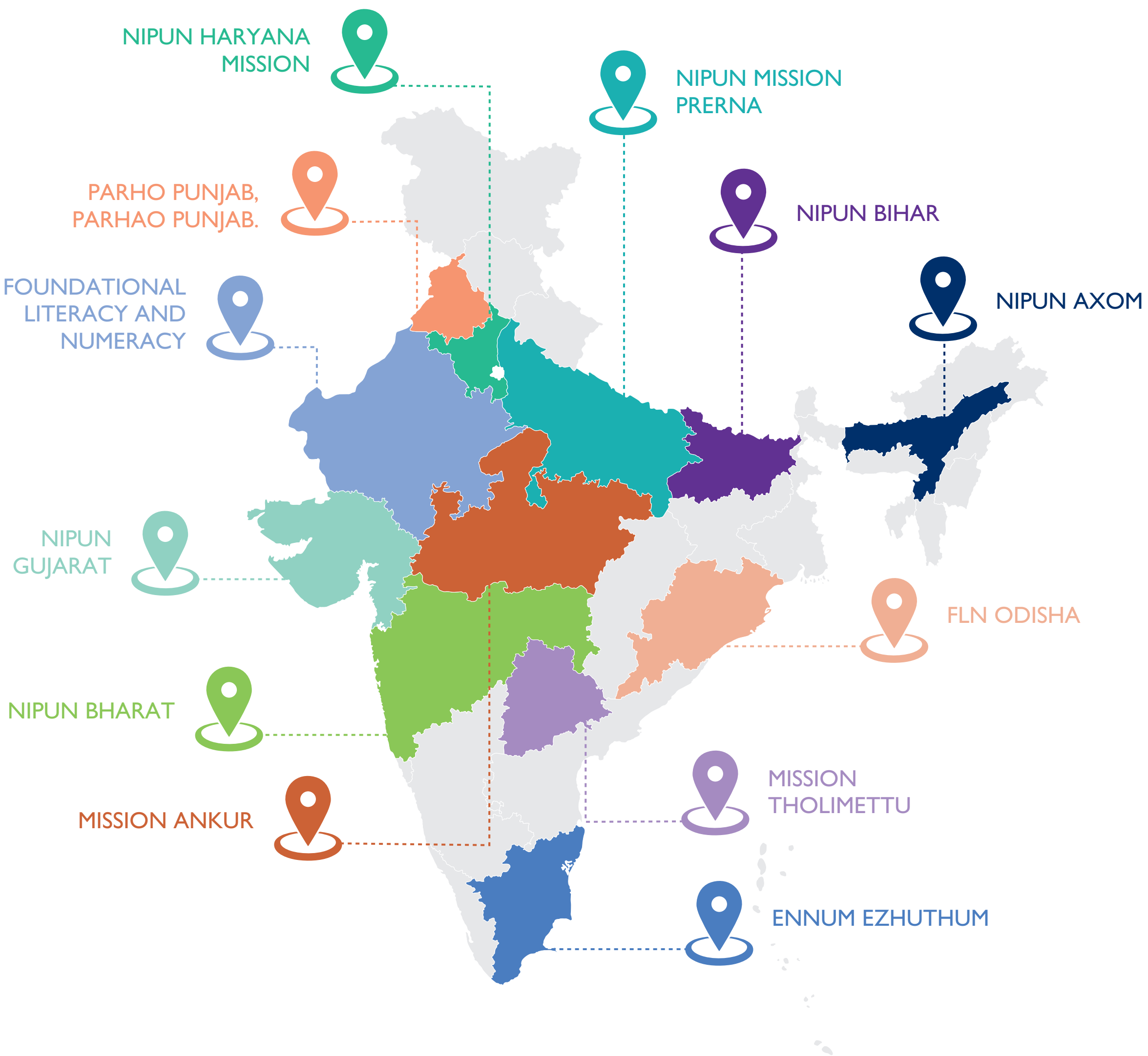


Figure 1 : State-Specific Names of NIPUN Missions Across India

1.2 Role of State-Wide Assessments in FLN Mission

The NIPUN Bharat Guidelines provide a comprehensive framework that delineates specific activities to be undertaken by each State and UT to ensure the successful implementation of the FLN mission. One of these crucial activities involves setting yearly targets aligned with the NIPUN goals or Lakshyas and monitoring annual progress accordingly by conducting periodic system-based assessments. This systematic approach helps to track advancements and ensures a focused effort in attaining the desired outcomes in foundational literacy and numeracy across the nation.

> "To know if the FLN Mission is meeting the expected learning outcomes and developing subsequent targets, student assessments will be carried out at least three times - baseline, midline, and endline. Evaluation would be conducted in a representative sample of schools that are selected systematically, using a complete sampling frame" — *Chapter 17: Need for Research, Evaluation and Documentation, Page 260, NIPUN Bharat Guidelines for Implementation.*

It is well known that student learning assessments play a crucial role in providing valuable information about students' achievement levels. When coupled with background surveys, these assessments can also offer insights into efficiency, access and equity in state education systems. Governments, international organizations, civil societies and other stakeholders in the education ecosystem increasingly acknowledge the significance of assessments for monitoring and enhancing student learning levels at the state level. As a result, there is a growing emphasis on developing robust systems for student assessment. Figure 2, illustrates how data gathered by measuring learning levels regularly, reliably and in a relevant manner through state-wide assessments can feed into a continuous improvement cycle for learning outcomes, highlighting the role of assessments in driving informed decision-making and instructional reform.

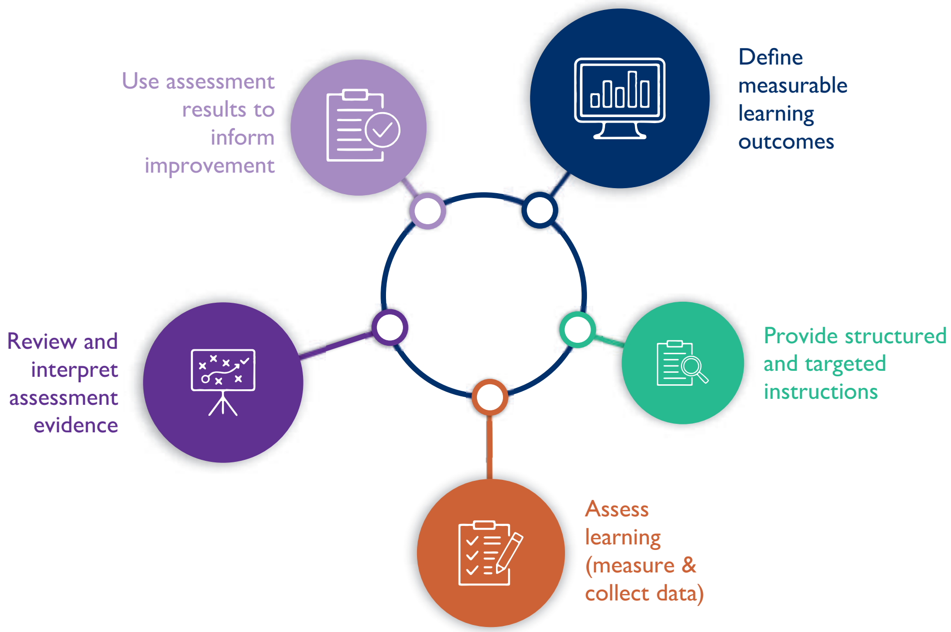


Figure 2 : Continuous Improvement Cycle for Learning Outcomes

1.3 Purpose of State-Wide FLN Assessments

Any large-scale system assessment must align with educational policy priorities and objectives. To be effective, these assessments should be specifically tailored to their intended purpose and clearly define how the collected data will be utilized. State FLN assessments can act as a system diagnostic where it serves three main purposes within the overall assessment architecture: diagnosis of educational needs, evidence-based intervention and monitoring of progress towards enhancing learning outcomes at the systemic level.

Diagnosis

Identify educational needs, pinpoint learning gaps / improvement areas by raising awareness using assessment findings.

Intervention

Plan evidence-based strategies to enhance teaching and learning materials, teacher training, and classroom management techniques.

Monitoring

Continuous monitoring to ensure progress in achieving FLN outcomes or provide information on where and how to intervene.

Figure 3 : Role of State FLN Assessments as a System Diagnostic

Based on the state's requirement and the broader purpose as mentioned above, one or more objectives and evaluation questions from the given list can be addressed using state-wide FLN assessments.

KEY OBJECTIVES

- Generate baseline data under the State FLN Program against which any future progress can be measured
- Set targets at State and District level under the State FLN Program against the benchmarks for reading and numeracy
- Use the data to guide the State & District level program planning, identify priority areas for intervention and inform design of TLMs, TPD programs
- Study achievement differences amongst various sub-groups, types of schools, etc.
- Study trend over time: baseline–midline–endline

EVALUATION QUESTIONS

1. What proportion of learners of Grades 1, 2 and 3 demonstrate proficiency in foundational literacy/language skills in the medium of instruction?
2. What proportion of learners of Grades 1, 2 and 3 demonstrate proficiency in foundational numeracy skills?
3. What school, teacher and household factors correlate with and impact learners' foundational literacy and numeracy outcomes?
4. What is the improvement in achievement of FLN skills over time?
5. What interventions can be planned in order to increase the learning levels?

Figure 4 : Mapping State Needs to FLN Assessment Objectives and Evaluation Questions

1.4 Importance of Assessing Early

As children grow older, acquiring literacy becomes more challenging. Struggles in school increase the risk of discouragement and dropout, depriving them of the potential lifelong benefits of education. Researchers like Karthik Muralidharan, Lance Pritchett and Luis Crouch emphasize the significance of foundational literacy and numeracy. Grade 3 is seen as a critical point for children to transition from "learning to read" to "reading to learn" in later grades. Failure to attain proficient reading skills by Grade 3 can lead to falling behind and a higher risk of dropping out.

Research also indicates that strong foundational skills have a positive impact on students' academic trajectory and school attendance *(Patrinos & Velez, 2009)*. However, if these skills are not acquired early, the learning gaps between students who have mastered foundational reading skills and those who have not tend to widen over time *(Darney, Reinke, Herman, Stormont & Ialongo, 2013; Scanlon, Gelzheiser, Vellutino, Schatschneider & Sweeney, 2008; Torgesen, 2002)*.

Additionally, research demonstrates that students who are fluent readers early on tend to improve their fluency further, while those struggling with reading maintain low levels of fluency.

Similarly, a strong foundation in numeracy during the early years is crucial for success in mathematics in the later years. Mathematical skills are in high demand in today's job market, as has been demonstrated by various economists. Most competitive jobs require some level of skill in mathematics. Mathematics fosters essential skills such as problem-solving, mental agility and cognitive flexibility, abilities that extend well beyond the classroom. These skills, developed through both formal instruction and informal experiences like street mathematics, are highly transferable to diverse areas of life and work. Hence, it becomes crucial to address gaps in foundational learning skills during early grades rather than waiting until later stages to tackle reading deficits at both individual and systemic levels.

To support the NIPUN Bharat Mission of achieving proficiency in foundational literacy and numeracy for every child by the end of Grade 3 by 2026-27, the FLN assessments should ideally be administered to students of Grades 1, 2, and 3 across both the areas — literacy and numeracy, against the mission's learning goals, known as Lakshyas.

This is crucial to gain an understanding of how many students are capable of reaching the learning objectives upon completing each grade. These assessments would reveal the present learning standards and pinpoint any disparities with the learning goals outlined by the Ministry of Education (MoE).

The Ministry of Education (MoE) defined certain learning goals in 2021, focusing on the highest level expected in foundational skills for each FLN grade. In 2024, they introduced additional competencies, including those that lead to the earlier-defined goals, to ensure a more comprehensive focus on the progressive development of skills across oral language, reading, writing and numeracy. The following figure and table list these goals respectively.

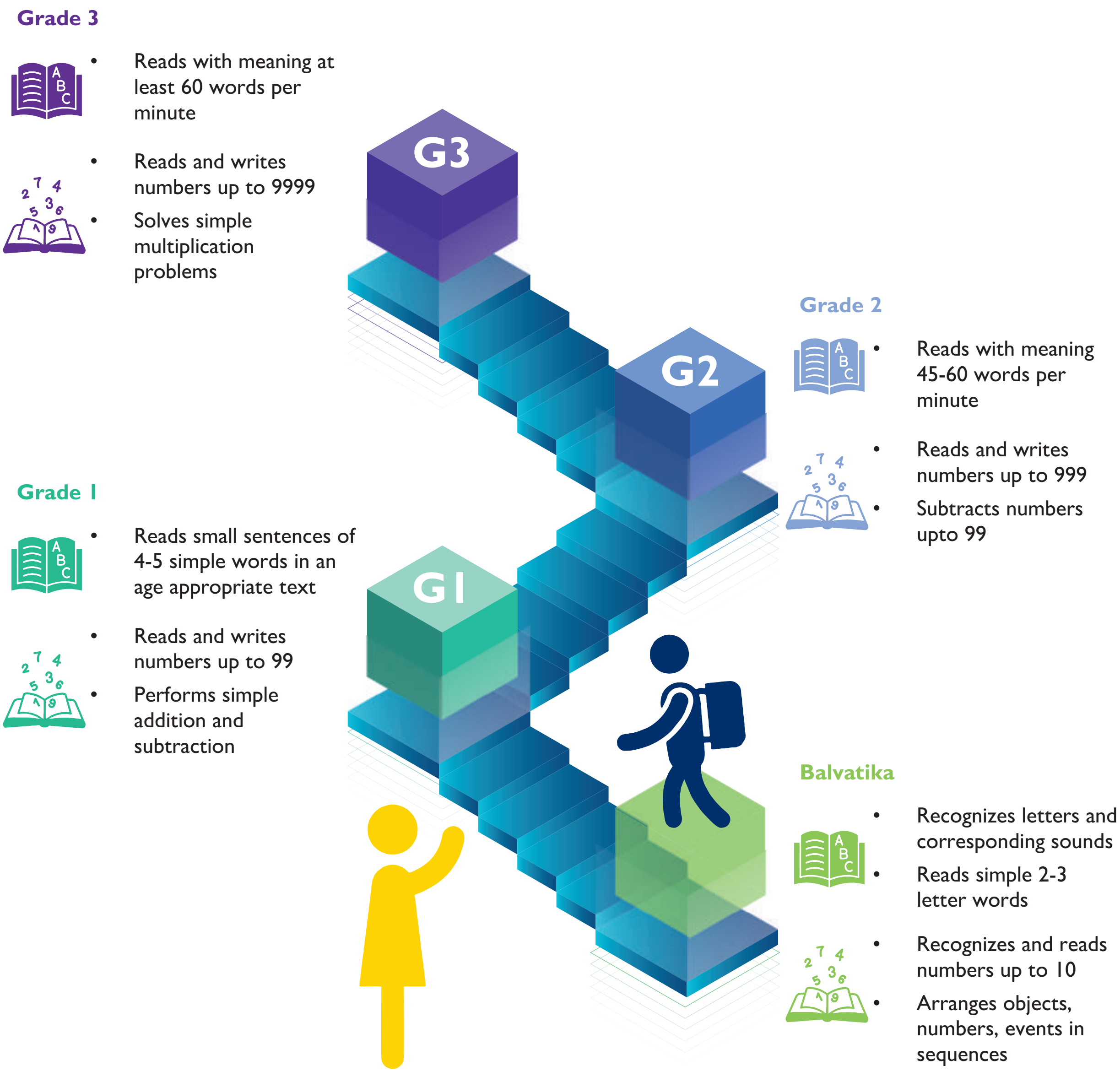


Figure 5 : 'Lakshyas' as given by NIPUN Bharat Mission. Source: NIPUN Bharat Guidelines MoE, 2021

Domain	MoE Description
Grade 1	
Oral Language	1. Talks about needs, familiar people and surroundings with peers and teachers.
	2. Discusses available print in the classroom, school and surroundings.
	3. Recites poems/songs with gestures.
Reading	4. Actively participates in story sessions, answers questions based on the story.
	5. Uses sound-symbol correspondence to read new words.
	6. Reads simple sentences in unfamiliar texts with accuracy.
Writing	7. Familiar with words with matras in known contexts (e.g., stories, poems).
	8. Writes simple sentences to represent meanings, objects or messages.
	9. Uses sound-symbol correspondence to write words with invented spelling.
	10. Writes 2-3 meaningful sentences with comprehension.
Numeracy	1. Counts objects and develops knowledge of numbers up to 20.
	2. Identifies and extends simple patterns in shapes and numbers.
	3. Uses addition and subtraction in daily life (sum ≤ 20).
	4. Observes and describes physical properties of 3D shapes.
	5. Estimates and verifies length using non-standard units (hand spans, cups, etc.).
	6. Identifies and sums coins and notes up to 20.
Grade 2	
Oral Language	1. Talks about print in immediate surroundings.
	2. Asks questions and listens to others during conversations.
	3. Recites songs/poems.
	4. Repeats familiar words from stories/poems/print.
Reading	5. Reads and narrates stories from children's literature/textbooks.
	6. Creates meaningful new words from given letters.
	7. Reads age-appropriate unfamiliar text fluently with understanding.
Writing	8. Writes small, simple sentences to express understanding and thoughts.
	9. Writes 4-5 small sentences legibly and meaningfully.
Numeracy	1. Counts up to 99, recognizes and compares numbers.
	2. Creates new patterns from shapes and numbers.
	3. Adds and subtracts numbers up to 99 in daily life.
	4. Uses repeated addition and division to understand multiplication and division.
	5. Identifies and describes 2D shapes.
	6. Estimates and compares length, volume and weight using non-standard units.
	7. Uses spatial vocabulary to describe position (near/far, inside/outside, etc.).
	8. Solves simple transactions using money up to 100.

Table 1 : 'Revised Lakshyas' as Given by MoE in 2024

2 DEVELOPING A WORK PLAN

2

Developing a Work Plan

2.1 Developing a Work Plan for Conducting the FLN Assessment

Developing a work plan at the outset of a project is crucial as it lays the foundation for successful execution. A well-crafted work plan ensures clarity, alignment and efficiency throughout the project lifecycle. It helps stakeholders understand the scope, objectives and steps involved in achieving project goals.

To create a work plan for a large-scale assessment cycle, begin by understanding the overarching objectives and desired outcomes of the assessment. Break down the assessment process into manageable tasks and establish clear timelines for each phase. Comprehending the various steps in the development of a large-scale assessment is essential for meticulous planning and seamless execution of any state-wide FLN assessment. A thorough understanding of each phase of the assessment cycle enables states to allocate resources strategically, set achievable timelines and ensure the completeness of essential components. This foresight fosters proactive decision-making, mitigates potential obstacles and optimizes the efficiency and effectiveness of the assessment process. Therefore, to formulate a comprehensive work plan—the first step in the nine-step assessment cycle covered in this chapter—it is essential to understand the steps illustrated in Figure 6, which are further elaborated upon in this playbook. Some of these steps or activities within them, can be carried out simultaneously, as illustrated in Figure 8 on Timeline and Key Activities. After completing the final step in the cycle, Dissemination & Use of Results, the process reiterates by returning to the first step to prepare for the following year's assessment cycle.

Figure 6: Steps in an Assessment Cycle

A work plan for conducting a large-scale assessment must include -

- Setting objectives
- Setting up a project team
- Setting project timelines
- Establishing a budget

2.1.1 Setting Objectives

A large-scale assessment will generate a significant amount of data about student learning levels. The objective of such an assessment must be clear to ensure that the assessment is appropriately designed and valid evidence is collected. Additionally, when an assessment is designed with a specific purpose in mind, it yields results that are most accurate and valid for that intended use. Aligning assessment design with its objective ensures that the inferences drawn are meaningful and appropriate. Broadly, there can be one or more of the following objectives of a large-scale assessment :

- Generate regular, reliable and relevant student learning outcome data under the State FLN Program against which any future progress can be measured
- Analyze trends in student learning over time by comparing data from the baseline, midline and endline assessments
- Investigate the differences in academic achievement among various population sub-groups, types of schools, and other relevant factors
- Use data to guide the state & district-level level program planning, identify priority areas for intervention, and inform the design of TLMs, TPD programs, and identify shifts in outcome indicators
- Understand the relationship between inputs and outputs through robust regression analysis utilizing background variables
- Establish targets at the state and district levels for the state FLN program, aligned with the benchmarks for reading and numeracy

During the initial planning phase, the state must establish a clear understanding of the objectives behind data collection and its subsequent utilization. To achieve this, the state can establish a steering committee comprising key government officials from the state education department, representatives from the central government, teachers, members of civil society and organizations partnering in assessment initiatives.

2.1.2 Setting Up a Project Team

Implementing the State FLN Assessment is a resource-intensive task that requires dedicated personnel to oversee the entire process. To administer and oversee the assessments, an 'Assessment Cell' should be established at the state/district level.

Ensuring state ownership of assessments is crucial for their effectiveness and sustainability. To achieve this, it is essential to involve key members from the state education department in various roles and responsibilities within the assessment team. This may include representatives responsible for policy formulation, curriculum development, assessment design, data analysis and program evaluation. By integrating individuals from the state education department into the assessment team, ownership and accountability are inherently established. These departmental representatives bring invaluable expertise, insights, and contextual understanding to the assessment process, ensuring that assessments align with state educational priorities, standards, and goals. Moreover, their involvement fosters a sense of ownership and commitment to the assessment initiative, leading to more meaningful and impactful outcomes for the state's education system.

The assessment cell at a state should comprise of the following members :

PROGRAM LEADER

Requirements

- A senior official with prior experience of managing and implementing state-level education surveys.

Responsibilities

- Prepare the work plan, manage the budget, and oversee the overall supervision of the state assessment programme.

NUMERACY EXPERT

Requirements

- They should have academic training on the structure and properties of early grade mathematics.

Responsibilities

- Contribute to the development of Numeracy assessment.

LANGUAGE EXPERT

Requirements

- Must have high proficiency and an academic training on the structure and properties of the language to be assessed.

Responsibilities

- Contribute to the development of test materials in the relevant language.

CURRICULUM EXPERTS

Requirements

- Should be knowledgeable about and responsible for the primary school curriculum.

Responsibilities

- Advise on how mathematics, reading and language are taught in schools.

LEAD DATA MANAGER

Requirements

- Must be familiar with computers to learn and understand the chosen data entry system and to assist in training others.

Responsibilities

- Oversee the processes of data entry, cleaning, management and ensuring quality control and accuracy.

STATISTICIAN

Requirements

- Should be knowledgeable in sampling and research design as well as statistics.

Responsibilities

- Carry out item analysis to check for item difficulty and item discrimination.

PSYCHOMETRICIAN

Requirements

- Degree in Psychometrics (PhD preferred) with proficiency in statistical software packages.

Responsibilities

- Design objective, valid, and reliable assessments and minimize measurement bias and ensure fairness in testing.

Figure 7: Composition of the Assessment Cell Team

These team members oversee the creation of an assessment framework, design of assessment tools, implementation of data collection processes, analysis of collected data and dissemination of results. To maximize human resources while working within budget constraints, team members may manage multiple assessment aspects. Additionally, outsourcing or temporary appointments to address specific aspects of the program can also be considered.

All the selected members of the assessment cell must possess certain attributes, including the commitment to being available as needed, the capacity to collaborate effectively throughout various stages of the assessment program, the willingness to acquire new skills and knowledge, the capability to handle multiple responsibilities within the assessment and the adeptness to operate within challenging time constraints.

2.1.3 Setting Up Timelines

Setting timelines early on during the planning phase of a state-wide assessment is paramount for ensuring smooth execution and timely completion of the entire assessment cycle. With the complete assessment cycle spanning 10-12 months, establishing timelines at the outset allows for effective coordination of activities, allocation of resources and monitoring of progress. It provides stakeholders with a clear roadmap, ensuring that each phase of the assessment is completed within the designated time frame, thereby facilitating informed decision-making and maximizing the impact of the assessment on education policy and practice.

Figure 8: Timeline and Key Activities Leading up to Results Dissemination

The illustration in Figure 8 depicts a range of activities scheduled before the dissemination of results, serving as an example

The timeline template promotes consistency and accountability among team members, ensuring everyone is aligned on project timelines and objectives. It also facilitates communication and collaboration, enabling effective coordination of efforts across various project phases. For effective use of time, many tasks can be carried out simultaneously, leveraging the expertise of different team members. For instance, the development of cognitive and contextual tools and sampling can progress hand-in-hand, maximizing efficiency and streamlining the overall project timeline.

2.1.4 Establishing a Budget

Conducting a State FLN Assessment involves several cost factors. Expenses can vary depending on the number of districts and the sample size of children assessed. Key cost drivers include material printing and the daily stipend paid to assessors, including their daily allowance for food and transportation during data collection. Additional costs arise from tool development workshops, assessor training, and the number of days required for workshops and assessments. A structured template, as shown below, can help organize and budget these expenses effectively.

Cost Headers	Particulars	No. of Units	No. of Days	Per day Cost (INR)	Total Cost (INR)	Actuals (INR)	Cost Borne by
1. Tool Development Workshop							
Personnel	Literacy expert (include all languages)						
	Numeracy expert						
	Background questionnaire						
	Typist						
	Trainer						
Material and Technology	Notebook						
	Pen						
	Data collection software subscription						
	Tool rendering personnel						
	Laptop / Computer for typing or tool rendering						
Logistics	Venue						
	Food						
	Lodging						
	Travel						
2. Field Pilot							
Personnel	State Chief Nodal Officer						
	State Nodal Officer						
	State Observers						
	District Coordinators						
	Block Coordinators						
	Assessors						

Cost Headers	Particulars	No. of Units	No. of Days	Per day Cost (INR)	Total Cost (INR)	Actuals (INR)	Cost Borne by
2. Field Pilot							
Material and Technology	Tool printouts						
	Physical manipulatives (if any)						
	Tablets for data recording						
	Blank papers						
Logistics	Food						
	Lodging						
	Travel						
3. Analysis of Pilot Data							
Personnel	IRT modeler and analyst						
	Inference/reporting expert						
Material and Technology	Software licenses						
4. Tool finalization workshop							
Personnel	Subject experts						
	Trainer						
Logistics	Venue						
	Food						
	Lodging						
	Travel						
5. Sampling							
Personnel	Sampling personnel						
Material and Technology	Sampling materials/software						
6. Training of Master Trainers							
Personnel	Main trainers						
	Master trainer participants						
Material and Technology	Print of tools						
	Projector rental						
	Miscellaneous printouts						
Logistics	Venue						
	Food						
	Lodging						
	Travel						

Cost Headers	Particulars	No. of Units	No. of Days	Per day Cost (INR)	Total Cost (INR)	Actuals (INR)	Cost Borne by
7. District Level Training of Assessors							
Personnel	Master trainers (facilitators)						
	Assessor participants						
Material and Technology	Print of tools						
	Projector rental						
	Miscellaneous printouts						
Logistics	Venue						
	Food						
	Lodging						
	Travel						
8. Data Collection							
Personnel	Assessors						
	State Chief Nodal Officer						
	State Nodal Officer						
	State Observers						
	District Coordinators						
	Block Coordinators						
Material and Technology	Print of tools						
	Blank sheets						
	Tablet						
Logistics	Food						
	Lodging						
	Travel						
9. Monitoring of Assessment Data							
Personnel	Data analyst						
Material and Technology	Software licenses						
10. Analysis of Assessment Data							
Personnel	Data analyst						
Material and Technology	Software licenses						

Cost Headers	Particulars	No. of Units	No. of Days	Per day Cost (INR)	Total Cost (INR)	Actuals (INR)	Cost Borne by
11. Development of Report							
Personnel	Report writers						
Personnel	Graphic designer						
Material and Technology	Design software subscription						
12. Printing & Dissemination of Report							
Material and Technology	Report printing						
Logistics	Courier/postage						
Logistics	Event setup (dissemination)						
TOTAL							

Table 2: Sample Budget Template for a Statewide FLN Assessment

3 DEVELOPMENT OF AN ASSESSMENT FRAMEWORK

3

Development of an Assessment Framework

3.1 Importance of an Assessment Framework

An assessment framework is a description of what an assessment intends to measure. It is an important document that ensures that the assessment is of high quality and provides useful information about the assessment design. Hence, the journey of assessment development always begins by developing a detailed assessment framework by the members of the assessment cell.

The purpose of creating an assessment framework consists of the following components -

- **Provides a Roadmap:** The framework acts as a guiding document for developing the assessment, ensuring that all important aspects of the assessment are covered.
- **Underpins Validity:** By making explicit what the purpose and objective of the assessment is, and what it will cover in terms of knowledge, skills and content, the framework ensures validity.
- **Enhances Consistency:** Developers can adhere to the framework resulting in consistent assessment development across time in case the assessment intends to measure trends in performance longitudinally.
- **Facilitates Communication:** It helps communicate the assessment's objectives and design to stakeholders, including educators and learners. It also helps to communicate what the assessment results mean.
- **Supports Improvement:** It allows for evaluation and refinement of the assessment over time to improve the effectiveness of the program.

The following figure shows the components of an assessment framework.

Figure 9: Components of an Assessment Framework

3.2 Proposed Framework for State FLN Assessments

For any system-wide assessment, the assessment must be standardized, that is, all test takers should be assessed using the same tool or a comparable tool and the process of administration and grading should be the same for everyone. It is also important to use an assessment that is reliable and valid.

Two such assessments for Foundational Learning that are being used across countries are Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessments (EGMA). EGRA and EGMA frameworks, developed by RTI International in 2006, serve as guides for developing contextualized assessment tools that offer straightforward measures of foundational literacy and numeracy skills at a low cost. Both EGRA and EGMA are one-to-one oral assessments designed to evaluate fundamental skills that form the basis of literacy and numeracy development in students from grades 1 to 3. These assessments have been implemented in at least 11 countries and 19 languages as part of the EdData II program and have been adopted in more than 30 countries including India across 60 languages. These tools are developed to assess the foundational learning levels, identify gaps in learning, inform educational policies and evaluate the effectiveness of interventions aimed at improving learning outcomes.

It is recommended to base the assessment framework for the system-wide FLN Assessment on EGRA and EGMA as these frameworks allow for efficient administration as they can be conducted within 20 minutes per child per subject, have high validity as they measure the targeted skill effectively, are reliable as they give consistent and dependable results, can be easily adapted for Indian languages, have a simple scoring system and allow for the creation of multiple versions. It not only assesses the targeted skill but also assesses the building skills required for the development of target skills.

The Assessment Cell's experts in literacy, numeracy and curriculum can use the following framework as a reference while developing frameworks for Grades 1 to 3, by appropriately scaffolding it for the lower grades. These frameworks include all building block skills for identifying learning gaps early. Skills applicable as per grade-wise curricular expectations should be selected and finalized for the assessment tool.

There are two types of tasks - timed and untimed. The purpose of untimed tasks is to check for accuracy while the purpose of timed tasks is to check for accuracy along with fluency.

Literacy Framework - Grade 3			
Task	**Skill**	**Nature**	**Items**
Oral Vocabulary (knowledge, comprehension, awareness of sound in spoken words)			
Listening Comprehension	Ability to answer questions based on a story read out by the assessor	Untimed	4
Oral Vocabulary	Ability to identify and name objects and actions	Untimed	10
Phonological Awareness	Ability to identify the initial or final sound of words into phonemes	Untimed	10
Reading (knowledge of the alphabet, letter-sound correspondence, automatic decoding, fluency, comprehension)			
Letter Naming	Identify letters presented in a random order	Untimed	15
		Timed	50
Word Reading	Read high-frequency grade level words presented in a random order	Untimed	15
		Timed	50
Non-word Reading	Read non-words composed of legal syllables presented in random order	Untimed	15
		Timed	50
Sentence Reading	Read small grade-level sentences	Untimed	10
Oral Reading Fluency	Read a narrative grade-level text with accuracy and fluency	Timed	10
Reading Comprehension	Answer direct and infer level questions based on the read text	Untimed	10
Writing (orthographic/phonological knowledge)			
Letter Dictation	Write dictated letters	Untimed	5
Word Dictation	Write dictated grade-level words	Untimed	5
Sentence Dictation	Write dictated grade-level sentences	Untimed	5

Table 3 : Literacy Framework for Grade 3, End of Grade Assessment

Numeracy framework- Grade 3			
Task	**Skill**	**Nature**	**Items**
Number Sense and Operations - Conceptual (Knowledge of number sense, place value, addition and subtraction facts)			
Number Identification	Recognize digits and numbers	Untimed	10
		Timed	20
Number Discrimination	Ability to make judgments about large and small by comparing quantities represented by numbers	Untimed	10
Counting in Bundles	Ability to apply the concept of place values in counting	Untimed	5
Number Patterns	Ability to discern and complete number patterns	Untimed	10
Addition Level 1	Ability to solve one-digit additions with automaticity	Untimed	10
		Timed	20
Subtraction Level 1	Ability to solve one-digit subtractions with automaticity	Untimed	10
		Timed	20
Number Sense and Operations- Procedural (application of number knowledge)			
Addition Level 2	Ability to use and apply procedural understanding on addition sums of two digits	Untimed	6
		Timed	12
Subtraction Level 2	Ability to use and apply procedural understanding on subtraction sums of two digits	Untimed	6
		Timed	12
Addition Level 3	Ability to use and apply procedural understanding on addition sums of three digits	Untimed	6
Subtraction Level 3	Ability to use and apply procedural understanding on subtraction sums of three digits	Untimed	6
Word Problems	Ability to apply addition and subtraction procedural skills to word problems	Untimed	6
Multiplication	Ability to apply the procedural knowledge of multiplying two numbers	Untimed	6

Table 4 : Numeracy Framework for Grade 3, End of Grade Assessment

All timed tasks are administered using a countdown timer, typically set to 1 minute. An exception is oral reading fluency (ORF) tasks, which may be timed for 1 or 2 minutes. While untimed tasks are not timed to the second, they are still time-limited, with nudge rules that are applied to ensure smooth progression. For example, in a word reading task, if a student is stuck on a word for 5–10 seconds, the assessor gently nudges the student to move on. The word is marked as incorrect.

Other than the above skills, counting fluency, one-to-one correspondence, measurement, shape recognition, division, data handling etc. can also be tested based on curricular importance.

3.3 Development of a Detailed Blueprint

Once the skills and content to be tested are identified, the next step is to create a detailed blueprint for the test developers to ensure content validity. The development of the test blueprint begins with a thorough textbook analysis of the grades for which the assessment is designed.

Some key steps in the textbook analysis for literacy are as follows :

- Creation of a list of the most frequent words from the stories / passages
- Identifying the most frequent letter and syllable combinations
- Segregating words according to the type, length, orthographic complexity, etc.
- Analyzing the passages to check for an average number of total words, sentence length, type-token ratio, lexical density, etc. That govern the text difficulty

During the textbook analysis for numeracy, it is important to analyze what kind of operations are introduced by which grade, when is the concept of zero introduced, what are the common types of word problems, what are the common contexts in the word problems, etc.

After a thorough textbook analysis, a detailed blueprint as per the grade level expectation must be created for each task. Following are examples of blueprints for literacy and numeracy.

Task: Word reading - Timed (Grade 2)	
Skill	Read high-frequency grade level words presented in a random order
Competency	The child should be able to read as many words as possible within one minute
Time	1 minute
Items	50
Item description	5 words with two-letters and no matra, 5 words with two-letters and one matra, 5 words with two-letters and two matra, 5 words with three-letters and no matra, 5 words with three-letter and one matra, 5 words with three-letter and two matra, 5 words with three-letter and three matra, 5 words with four-letter and no matra, 5 words with four-letter and one matra, 5 words with four-letters and two matra. * All the above words are to be arranged in a random order.
Administration	One on one - verbal. The child will be shown a list of all 50 words and asked to place their finger on each word and read them one by one quickly. As soon as the child starts reading, the timer for 1 minute starts. Nudge rule - If a child is stuck at a word for more than 3 seconds, they are asked to read ahead leaving the word on which they are stuck Stop rule - If a child reads four consecutive words incorrectly, the timer is stopped and the task ends.
Grading	0 = no response / incorrect 1 = correct
Data analysis	Fluency rate [(words read correctly / time taken in seconds)*60]

Table 5 : Item Writing Template for Literacy

Task: Addition Level 2 - Untimed (Grade 2)	
Skill	Ability to use and apply procedural understanding on addition sums of two digits
Competency	The child should be able to perform the addition of two-digit numbers up to 99
Time	Untimed (but maximum up to 3 minutes)
Items	6
Item description	Addition of two-digit with two-digit without carry forward sum less than 50 Addition of two-digit with two-digit with carry forward but sum less than 50 Addition of two-digit with two-digit without carry forward sum less than 99 Addition of two-digit with two -digit with carry forward sum less than 99 Addition of two-digit with single-digit without carry forward sum less than 50 Addition of two-digit with single-digit with carry forward sum less than 50 * All the above sums are to be arranged in a random order.
Administration	One on one - written. The child will be shown the list of all 6 sums and asked to solve the sums one by one. Nudge rule - If a child is stuck at a sum for more than 5 seconds, they are asked to move ahead leaving the sum on which they are stuck. Stop rule - If a child solves three consecutive sums incorrectly, the task is stopped.
Grading	0 = no response / incorrect 1 = correct
Data analysis	Accuracy [(number of sums solved correctly / total number of sums]

Table 6 : Item Writing Template for Numeracy

There are two types of stop rules :

1. **Intra-task rule / Stop rule -** That applies within the task (for example, if a child reads four consecutive words incorrectly, the task ends).
2. **Inter-task rule / Skip rule -** That applies across tasks (for example, if a child cannot read at least four words, they will not be asked to do sentence reading or ORF tasks).

These stop rules make the test adaptive as per the child's learning level and also limit the time taken for administration.

3.4 Guidelines for Task/Item Design

Along with a detailed blueprint, it is also important to create guidelines for item design. These guidelines provide clear and detailed instructions for developing each type of task, ensuring consistency and alignment with the assessment framework.

Below are two examples. The remaining guidelines are provided in Appendix 1 and 2.

LISTENING COMPREHENSION

The child has to listen to a story read out by the assessor and then orally answer 4 comprehension questions based on the same story.

TASK DEVELOPMENT

Story

- Create third person narrative of a story with 1/2 characters with only 1 proper noun.
- Make original or new story —not similar to common stories or legends with which children maybe familiar.
- Have a logical flow with a beginning, a problem/confrontation and a resolution to it.
- Do not include a story title or illustrations.
- Avoid violence or other themes that might be troubling to children.
- Use positive gender representations.
- The entire story should be in the same tense.
- The length of the story should be grade appropriate.
- The words used should be grade appropriate in terms of frequency, length, number of syllables.
- Have grade appropriate grammar and sentence structure (avoid conjunctions and dependent clauses).
- Avoid including dialogue with quotation marks.

Questions

- Three should be explicit (answer found directly in text).
- One should be inferential (answer not found directly in the story).
- The questions should refer to the story, but do not need to be distributed evenly or aligned to sequential ideas.
- Include the correct answers next to the questions for easy reference for assessors.

Figure 10: Example of a Task Development Guideline for a Literacy Task

PRE NUMERACY SKILLS- PRE-MEASUREMENT

The child is able to apply the knowledge of pre numeracy skills (length, weight, capacity) in different situations in an untimed task (7 questions).

TASK DEVELOPMENT

- Develop 7 items aligned to grade-appropriate learning outcomes related to measurement (length, weight, and volume).
- Use curriculum-specific terminology and contexts familiar to children at each grade level (Grades 1–3).
- Design tasks that involve: Comparing 2 to 3 objects based on a single measurable attribute (e.g., longer/shorter, heavier/lighter, holds more/less). Ordering 3 to 4 objects by one attribute (e.g., weight or length).
- Label the objects clearly as A, B, C (and D, if needed) to help children refer to them easily during comparisons or ordering.
- For each item, mark the correct answer clearly for FIs (Field Investigators) to use during grading.
- Ensure that objects shown are visually distinct and clearly differ in the attribute being measured.
- For Grade 3, include estimation tasks using non-standard units (e.g., hand spans, cups, blocks), if included in the curriculum.
- Use clear, proportionally accurate and age-appropriate images to support visual comparison.
- Choose objects that are culturally relevant and familiar to children in the local context.

Figure 11: Example of a Task Development Guideline for a Numeracy Task

4 DEVELOPMENT/ ADAPTATION OF COGNITIVE TOOLS

4

Development/Adaptation of Cognitive Tools

4.1 Organizing a Tool Development Workshop

The development of FLN assessment tools in a state should involve a collaborative effort from a diverse group of stakeholders such as SCERT/DIETs/SRGs, government staff, teacher trainers, former or current teachers, language experts from curriculum development centers and local colleges and universities, representatives from non-governmental organizations (NGOs) and international agencies and members of the Assessment Cell. Including a wide range of perspectives can help ensure the tools are adapted effectively to the local context, language and curriculum. The development of both cognitive and contextual tools takes place in a workshop. Contextual tools have been discussed further in Chapter 5.

The tool development workshop should ideally be held at least 4–6 months before the date when the assessment is planned with students. Planning for the workshop should begin well in advance, including finalizing the number of members from the State Resource Group (SRG) required and ensuring all preparatory materials are ready. At least three resource persons per subject should be present for item development during the workshop. If assessments are to be conducted in multiple languages, a minimum of three experts per language must be included for the development of literacy tools.

Sample schedule of a workshop for developing FLN assessment tools -

Sessions	Broader Agenda	Time	Resources Required
Day 1			
Context Setting for Workshop	• Setting context on importance of state wide EGRA-EGMA based assessment • Assessment plan (scale, dates etc.) • Sharing of agenda and expectations from the workshop • Explanation of tool development process	1 hour	PPT
Introduction to FLN Framework	• Sharing assessment framework for all subjects • Discussion on construct definitions and expectations from children • Walkthrough of entire toolkit (glimpse of assessment app, stimulus sheet and worksheets) • Grouping of participants as per expertise	1.5 hours	Assessment framework and construct definitions
Lunch break			
Working in two groups - Language and Numeracy			
Introduction to the Assessment Tasks	Walkthrough on language and numeracy tasks in respective groups (instructions plus blueprint)	2 hours	Language and numeracy tool templates
Item Writing	Explanation of guidelines followed up with item writing • Letter naming • Word reading • Non word reading • Measurement • Shapes • Money	2 hours	Language and numeracy tool templates and guidelines
Day 2			
Item Writing	Explanation of guidelines followed up with item and instruction writing • Listening comprehension • Oral Vocabulary • Phonological awareness • Number identification • Number comparison • Number Pattern	3 hours	Language and numeracy tool templates and guidelines

Sessions	Broader Agenda	Time	Resources Required
Day 2			
Tea break			
Item Writing	Explanation of guidelines followed up with item writing • Sentence reading • ORF and RC • Dictation • Addition level I and 2 • Subtraction level I and 2 • Word problems	3 hours	Language and numeracy tool templates and guidelines
Lunch break			
Peer Review in Subject Groups	Peer Review and corrections [items plus instructions]	4 hours	Language and numeracy tool templates and guidelines
Day 3			
Item Panelling Across All Subjects	Panelling of all the tasks and instructions across all subjects	3 hours	Language and numeracy tool - word docs for tasks, stimulus sheet and worksheets
Tea break			
Review of Final Print		4 hours	Language and numeracy tool - PDF docs for tasks, stimulus sheet and worksheets

Figure 12 : Sample Item Development Workshop Schedule

4.1.1 Materials Needed For Tool Development Workshop

Organizing a workshop for the development and/or adaptation of FLN assessment tools requires a variety of materials to facilitate effective collaboration, communication and learning among participants.

1. **Workshop Agenda**
A clear and detailed agenda outlining the workshop objectives, sessions

2. **Participant List and Contact Information**
A list of participants, their affiliations and contact information

3. **Background Documents**
Provide relevant background information on EGRA and EGMA-based tools

4. **Curriculum and Learning Outcomes**
Access to the state's curriculum and learning outcomes for early grade reading and mathematics

5. **EGRA and EGMA Manuals and Templates**
Existing manuals and templates for EGRA and EGMA, which serve as a starting point for development

6. **Language and Cultural Resources**
Materials related to the local language, such as dictionaries and culturally relevant texts

7. **Presentation Materials**
Slide decks, flip charts, whiteboards and markers

8. **Stationery**
Provide participants with notepads, pens, sticky notes

9. **Computers and Internet Access**
Internet access can help participants to research, collaborate and create digital resources

10. **Breakout Session Materials**
Materials such as flip charts, markers, and sticky notes to facilitate group discussions and collaboration

11. **Evaluation Forms**
Evaluation forms help to gather feedback on the workshop and identify areas for improvement

12. **Refreshments and Meals**
Participants should have access to refreshments, water and meals during the workshop

Figure 13 : Material Required for Item Development Workshop

4.2 Importance of Tool Adaptation Process

Once the assessment framework is finalized and a detailed blueprint for tool development is ready, the process of task development should also be undertaken by the team of literacy, numeracy and curricular experts and other item developers from the assessment cell. Along with the EGRA and EGMA frameworks the tasks in the EGRA and EGMA tools have also been rigorously tested internationally and established as a reliable measurement for foundational literacy and numeracy. Since EGRA has been used in countries that have English as the primary language, it has been observed that directly adopting EGRA tools does not work for the Indian languages which are phonetic, i.e. each *'akshara'* can be pronounced independently. Similarly, there is a difference in the numeracy curriculum for FLN grades in India and other countries. Given this aspect, rather than using them directly in their existing form, EGRA and EGMA should be used as a guide that can be adapted, to accommodate the unique attributes of the language being assessed as well as the state curriculum.

Figure 14: Importance of Conducting Item Development Workshop

4.3 Steps Involved in Adapting FLN Assessment Tools

The FLN assessment tools need to be adapted to the specific context and the medium of instruction of a given state before they can be used to effectively measure foundational skills. The general steps involved in developing the assessment tools for a state are as follows:

1. Identification of domain, tasks and skills as per the curriculum expectation

The first step towards the development of an assessment is to create an assessment framework that identifies the specific tasks that need to be assessed based on the state's curriculum and learning outcomes. Before the workshop, language experts from the Assessment Cell should review the framework and guide the appropriateness of each task, considering how the construction might be affected by language properties. They may choose to add or delete tasks to better fit the language and context. Similarly, numeracy specialists should review the numeracy tasks, suggesting which tasks to retain or delete from the framework.

2. Development of a detailed blueprint

Using the guidelines provided by EGRA-EGMA, the next step after finalizing the framework is to create a detailed blueprint for each task to be developed. For this, the team must refer to the textbooks and any other teaching learning material used in the state. This has been discussed in detail in Chapter 3.

3. Direct translation of standardized instructions for administration

The instructions for the assessors as well as the instructions given to the child must be directly translated into the medium of instruction in which the assessment is to be conducted. A direct translation is preferred for the instructions as these instructions are widely tested for reliability.

4. Language and context adaptation

Since languages have unique structures, phonemes, and orthographies that can impact the way students acquire FLN skills, merely translating the tools without considering these linguistic differences can lead to assessment items that are not appropriate or meaningful for students in the target language. Adaptation also includes ensuring that the content is culturally appropriate and selecting relevant examples and illustrations. It is essential to work closely with local experts and educators during this process to ensure accurate adaptation.

> For example, translating a syllable-segmenting task from English to Hindi when the word being segmented is "yesterday" (a three-syllable word) would result in comparing it to 'कल' (a two-syllable word). This would reduce the reliability of the assessment instrument and the validity of the cross-linguistic comparisons of the result. Therefore, it is important to adapt the tasks and not translate them. Adaptation would ensure that the tasks are of equal difficulty and pertain to the grade level in each context.

5. Development of assessment items

The tasks need to be accurately designed to assess students' proficiency levels and be aligned to the state's curriculum and learning objectives. Involving local educators and subject matter experts can help ensure that the assessment items are valid. A detailed textbook analysis exercise must be undertaken to create a list of frequently used words in the textbooks, along with an analysis of the context and vocabulary used in the reading passages for the literacy tools. Similarly, a detailed analysis of numeracy textbooks must also be undertaken to map the curricular limitations. The tasks must be developed adhering to the guidelines provided in Appendix I and 2.

6. Development of assessment toolkit

Once the assessment tasks are developed, they must be put together in the form of an assessment kit. This kit would include a stimulus sheet (a print-ready material that presents the list of letters, words, non-words, passages, number recognition task, etc. that the child has to read from), a response sheet (a print-ready material where the child solves a task, e.g., dictation, number operations, etc.) and data collection forms (preferably a digital application).

7. Reviewing and finalization of the draft toolkit

A review team must use validity criteria to suggest changes to the draft assessment tools. A checklist can be used to ensure that all suggestions provided in the guidelines are followed. Once the tools are reviewed and finalized, they must be piloted in the field. The process of field pilot is discussed in Chapter 6.

Case Study:

Developing a Large-Scale FLN Assessment for a Multilingual State

Context and Objective

In an effort to strengthen foundational literacy and numeracy, a large-scale assessment was designed to capture students' learning outcomes across three mediums of instruction in a state. The primary goal was to ensure the assessment tools were comparable across languages while remaining contextually appropriate and aligned with the state's curriculum.

Designing the Literacy Tool

The literacy assessment was built using a common framework and blueprint for all three languages. However, to maintain linguistic and contextual relevance, items were developed separately for each language. A detailed textbook analysis formed the foundation of tool development, ensuring that reading tasks reflected the vocabulary and structures familiar to students.

High-frequency words from the textbooks were prioritised in reading and dictation tasks such as word reading, sentence reading, oral reading fluency (ORF) and dictation of words and sentences. Frequency of some words, letters and matras found in the Grade 2 textbook (Bengali medium) of this state's Bengali textbook was as follows.

Words	Frequency
করে	49
সে	35
কী	28
একটি	25
সে	23
কী	23
একটি	21
সে	20
আর	17
থেকে	17
যে	14
খেতে	14
একটি	12
কেন	11
তার	11
সে	11
আছে	11
করে	11
লাগল	10
তাদের	10
খুব	10
করল	10
সবাই	10
থেকে	10
আর	9
এসো	8

Matras	Frequency
া	871
ে	599
ি	409
ু	187
ো	148
্	90
ী	37
ৌ	19
ূ	17
ৈ	16
ু	15
ৃ	15

Letters	Frequency
র	784
ন	504
ল	472
ক	471
ব	427
ত	361
দ	306
ম	297
প	288
স	266
য়	193
গ	190
ট	171
ছ	166
জ	150
শ	143
য	131
খ	128
চ	124
ড	119

Additionally, various text features from the textbooks were reviewed to guide the development of listening and reading comprehension passages. These features included average number of words, number of sentences per paragraph, average sentence length, and type-token ratio (TTR) which is a measure of lexical diversity. A TTR value closer to 1 indicates a greater proportion of unique words, which generally signals higher text complexity. However, in Grades 1 and 2, textbook stories tend to use repetitive vocabulary to build familiarity and reinforce word recognition in context.

For example, stories in the Grade 2 language textbook typically consisted of paragraphs with 60–70 words, spread across 8–9 short sentences, each averaging 7–8 words and a TTR of around 0.6 to 0.7. These patterns served as reference points while developing reading comprehension passages. For listening comprehension, stories from Grade 1 textbooks were analyzed, as the task was intended for students at a lower proficiency level and consisted of 40-45 words.

An empanelling exercise was conducted with state resource groups to rigorously review and validate the tools, ensuring consistency in difficulty levels and cognitive demand across languages.

Following tool development, Item Response Theory (IRT) analysis was conducted on pilot data to statistically validate tool comparability across the three languages. This was followed with making final changes in all the three language tools ensuring that the reflections from IRT findings were appropriately incorporated in all of them.

Designing the Numeracy Tool

For numeracy, a common assessment tool was developed. However, careful contextualisation and translation of instructions were undertaken to cater to the linguistic diversity of the state:

Two of the mediums shared numerals with significant similarities, allowing for a smoother translation process.

However the third medium uses English numerals, requiring careful alignment to ensure comparability in both presentation and cognitive demand.

The grading instruction ensured that response in any language was considered correct in the numeracy tool.

Designing the English Tool

The English assessment remained common across the state. However, to support the field investigators (Fis), instructions were translated into all three mediums of instruction, similar to the approach adopted for the literacy and numeracy tools.

Key Enablers of Tool Comparability

- Textbook Analysis: Provided critical insights into grade-level vocabulary and concepts across languages.
- Common Blueprint: Ensured uniformity in structure, content areas, and competencies assessed.
- Empanelling Process: Enabled thorough expert review and standardisation across languages.
- Pilot Testing and IRT Analysis: Strengthened the psychometric robustness and comparability of tools.

Conclusion

The multilingual assessment design demonstrates a successful model of maintaining comparability across diverse linguistic contexts while respecting language-specific nuances. Through careful planning, expert validation, and rigorous piloting, the assessment tools were able to provide reliable, comparable data on foundational learning across all three mediums of instruction.

4.4 Maintaining The Rigor of FLN Assessment Tools

The adaptation of FLN assessment tools for specific states and languages is essential for ensuring that the assessments are linguistically and culturally relevant, aligned with local curricula and learning outcomes, valid and reliable, engaging for students, fair and equitable, and capable of generating context-specific insights.

However, while developing the assessment tools, the assessment cell must maintain the following characteristics of the tool:

- The tool must be quick to administer (under 20 minutes).
- The tasks must have high validity (measures what it is supposed to measure).
- Adapt instructions to match local educational practices and norms, making them meaningful and relatable.
- The tool must allow for the development of parallel forms (versions). Ideally, these parallel forms that can be used across years should be developed in one go.
- The task administration instructions should be standardized so that they result in reliable and valid data.
- Ensure instructions are culturally relevant and avoid confusing or offensive references.
- Maintain consistent language and terminology throughout the instructions and assessment items to avoid confusion.
- The tasks must be administered orally in a language the child can understand - preferably home language.

- Ensure translations accurately convey the original meaning, accounting for language-specific nuances and expressions and use clear and simple language for easy understanding by teachers, assessors and students.
- The tasks must be easy to score and follow standardised scoring.

4.5 Key Components of FLN Assessment Toolkit

The entire FLN assessment toolkit comprises three components:

i. Assessment forms deployed on a digital app, which not only captures data but also provides standardized instructions for both assessors and students.

ii. A copy of the stimulus sheet which comprises the tasks that students are required to look at, read, and respond to. For example, in literacy - the passage which the students have to read for checking oral reading fluency, in numeracy - the number comparison tasks.

iii. Worksheets for students to solve numeracy tasks like addition, subtraction, word problems and write words and sentences dictated to them.

Data Collection App | **Stimulus Sheet** | **Worksheet**

Figure 15: Components of FLN Assessment Toolkit

4.6 Using Data Collection Application

Data collection applications like Tangerine and SurveyCTO are widely used for efficient data collection in various fields, including education and research. However, while developing the assessment tools, the assessment cell must maintain the following characteristics of the tool:

- **Tangerine Developed by RTI** is specifically designed for foundational literacy and numeracy assessments. It allows assessors to record student responses electronically on tablets or mobile devices, eliminating manual data entry and minimizing errors. Tangerine is particularly useful in low-resource settings where traditional paper-based assessments may be challenging.

- **SurveyCTO Developed by Dobility Inc** is a versatile data collection platform used for a wide range of surveys, including education, health and socio-economic research. It supports offline data collection, integrates well with cloud-based storage and offers advanced security features to protect sensitive information. With tools for real-time monitoring and validation, SurveyCTO enhances data accuracy and efficiency in large-scale studies.

Both applications improve data reliability, reduce errors and streamline the assessment process, making them valuable tools for large-scale research and educational evaluations. The following figure shows the features to consider for selecting any data collection app for FLN assessments.

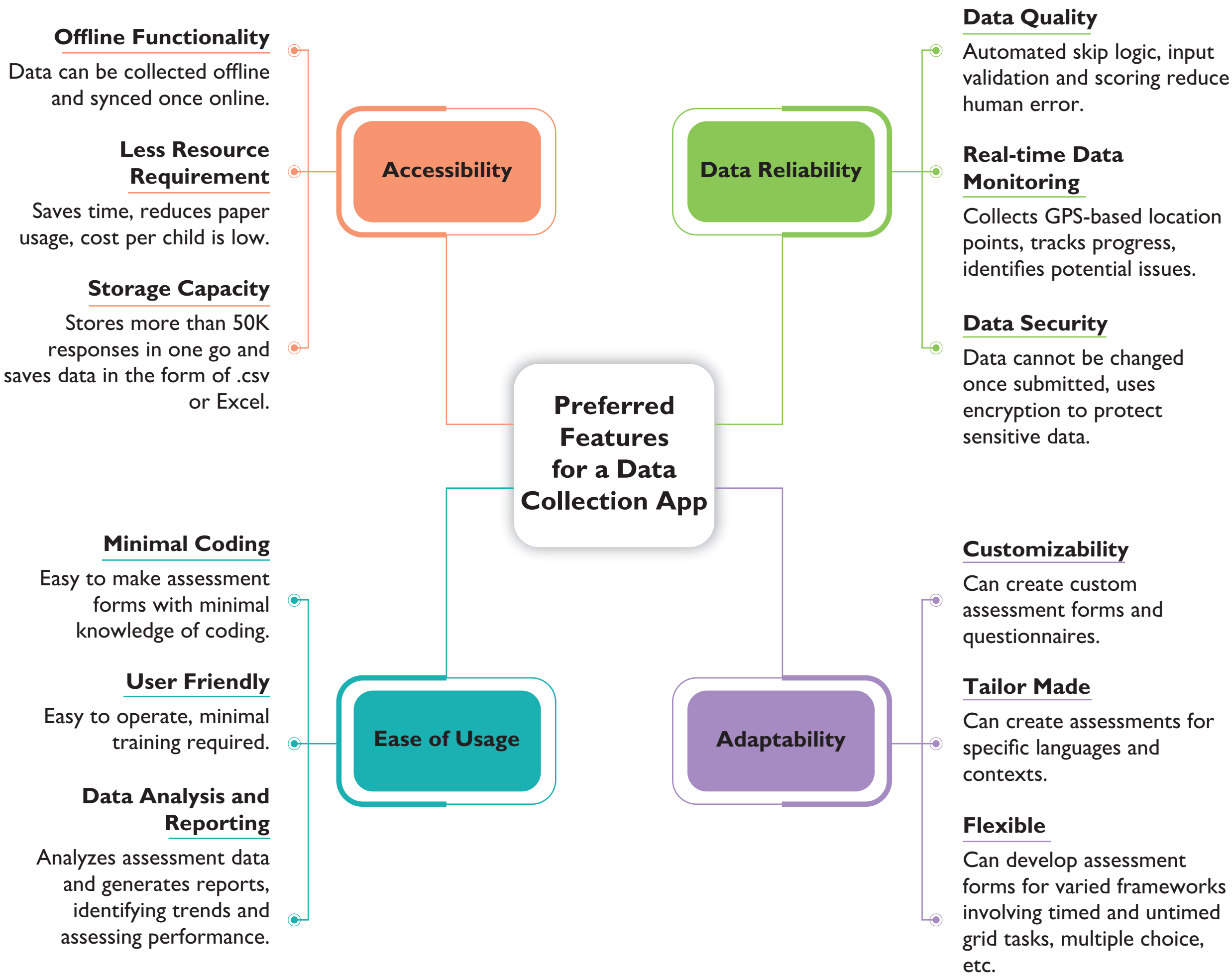


Figure 16: Preferred Features for a Data Collection App

5 DEVELOPMENT/ ADAPTATION OF CONTEXTUAL TOOLS

5

Development/Adaptation of Contextual Tools

5.1 Importance of Background Questionnaires

The workshop for item development, facilitated by experts, should also focus on the development of supplementary research instruments or background questionnaires to be administered alongside the main assessment tools. Background questionnaires serve several important purposes in a student assessment for foundational skills. They are crucial in understanding the broader context of a student's learning environment, which goes beyond academic abilities. These questionnaires collect data on various factors such as socioeconomic status, parental education, school infrastructure and community perceptions, which can help in the following ways:

1. **Contextualize Assessment Results:** Background questionnaires provide valuable contextual information that helps researchers and policymakers understand the factors that may influence students' performance on foundational skill assessments. This context can be crucial for interpreting the results and identifying areas where interventions may be needed.

2. **Identify Factors Affecting Performance:** By collecting data on students' demographic characteristics, socioeconomic status, home environment, and school experiences, background questionnaires can help identify factors that may be associated with students' performance in foundational skills. This information can be used to inform targeted interventions and support policies aimed at addressing specific challenges and improving educational outcomes. This also enables educators and policymakers to design more targeted and effective strategies.

3. **Monitor Equity and Inclusiveness:** Data collected through background questionnaires can help monitor equity and inclusiveness in education by examining disparities in students' performance based on factors such as gender, socioeconomic status, ethnicity and location. This data can guide policies and interventions designed to ensure equitable access to quality education for all students. For example, students from marginalized communities frequently face limited access to quality education, which contributes to higher dropout rates.

Background questionnaires can bring these disparities to light, allowing for targeted actions to address them.

4. **Evaluate the Impact of Interventions:** Background questionnaires can serve as a valuable source of data for evaluating the impact of educational interventions and programs on students' foundational skills. By comparing the information collected before and after an intervention, researchers and policymakers can assess its effectiveness and make data-driven decisions about resource allocation and future initiatives.

5. **Give a Holistic Understanding of Student Learning:** Testing students solely on academic concepts is insufficient because their environment, access to quality education, and community perceptions also impact their performance. Background questionnaires help gather information on these external factors, providing a more comprehensive understanding of a student's learning conditions.

6. **Provide Information on the Support Ecosystem:** The data from background questionnaires can reveal the level of support students receive from their families, teachers and communities. A lack of support can negatively impact a student's engagement with school education, which will be reflected in their assessment performance.

5.2 Factors That Influence Learning

1. Teacher Influence and Pedagogical Approaches

Teachers are pivotal in shaping the learning outcomes of students. The teaching pedagogy they use, their subject expertise and their ability to connect with students are critical factors. Effective teaching methods should be adaptive, using a mix of instructional strategies to meet diverse student needs. Teachers must leverage visual aids, interactive activities and discussions to cater to different learning styles, thereby enhancing student engagement and understanding.

In assessments, the quality of teaching is often evaluated through teacher-student interactions, instructional clarity and the ability to foster critical thinking. Teachers also play a crucial role in motivating students, both intrinsically and extrinsically, which can have a significant impact on students' performance in assessments.

2. Student Characteristics

Students' personal interests, curiosity and external factors like recognition and rewards directly influence their learning effectiveness. A student's prior knowledge and life experiences also play a significant role in their ability to grasp new concepts. Assessments often examine how students apply their knowledge and skills in various contexts, making prior learning a crucial element for success. Additionally, students' social and emotional well-being, including their self-efficacy and resilience, are critical. These aspects are sometimes reflected in assessments, which may include measures of students' attitudes toward learning, self-regulation and persistence.

3. School Environment and Resources

The learning environment extends beyond the physical classroom to include the overall school climate, access to learning resources and support systems. A safe, comfortable and well-resourced school environment is fundamental to effective learning. This includes not only physical resources like textbooks and technology but also access to qualified teachers, extracurricular activities and support services.

In assessments, the school environment is often assessed in terms of resource availability, teacher quality and school leadership. These factors collectively contribute to creating an environment that is conducive to learning and helps in achieving higher educational standards.

4. Parental and Community Support

Parents and the broader community significantly influence a child's learning journey. Parental involvement in education, such as monitoring academic progress, encouraging learning at home, and supporting educational goals, is closely linked to student achievement. Community resources, such as libraries, cultural institutions and after-school programs further enrich the learning experience.

Assessments sometimes consider parental engagement and socio-economic factors when evaluating educational outcomes. The support from parents and the community can provide students with additional learning opportunities and reinforce the knowledge and skills gained in school.

The following figure illustrates the different types of contextual questionnaires—student, parent, teacher and school—and the various factors within each that can influence student learning.

Student Questionnaire

- Age
- Gender
- Social Category

Parent Questionnaire

- Educational Qualifications
- Occupation
- Language
- Home Environment Details
- Socio-economic status- Household Details
- Student's Attendance

Teacher Questionnaire and Learning Process

- Personal Detail
- Educational and Professional Qualifications
- Experience
- Training
- Language of Instruction
- Use of Learning Material like Library, Labs, etc.
- Lesson Planning
- Classroom Composition and Monitoring
- Attendance

Teacher Questionnaire and Learning Process

- UDISE Code and School Information
- Head Teacher Qualifications and Experience
- School Management Committee
- Infrastructure Details
- Observation Form

Figure 17: Types of Contextual Questionnaires and the Factors that Affect Student Learning

6 PILOTING

6

Piloting

6.1 What is a Pilot Study?

In large-scale assessments (LSAs) of student learning, a pilot study is a small-scale preliminary test conducted before the full assessment. It helps refine instruments, procedures, logistics and analysis methods to ensure reliability and validity. Rather than producing definitive data, pilot studies identify potential issues in test items, administration protocols, data collection and scoring. Field trials, often used interchangeably with pilot studies, may also assess item alignment with the assessment framework, the functionality of digital platforms and training for test administrators.

The table below summarizes the key uses and importance of pilot studies in large-scale assessments by categorizing their impact on assessment quality, logistics, stakeholder engagement, risk management and capacity building. It highlights how pilot studies help refine test items, improve logistics, ensure fairness and enhance overall assessment effectiveness.

Uses and Importance of Pilot Studies in Large-Scale Assessments		
Assessment Quality & Validity	Item Reliability & Validation	Ensures test items align with assessment objectives, using IRT or CTT to flag items that are not performing well.
	Evaluation of Instruments	Reviews test content for clarity, cultural appropriateness and technical accuracy. Identifies typographical and translation errors.
	Ensuring Fairness & Equity	Detects biases and ensures all students have equal opportunities to demonstrate their skills.
Operational Feasibility & Logistics	Testing Logistics	Assesses implementation challenges, including costs, materials, invigilator availability and technology infrastructure.
	Data Quality & Feasibility	Evaluates the completeness and accuracy of responses, tests cleaning and scoring procedures while ensuring ethical compliance and privacy regulations.

Uses and Importance of Pilot Studies in Large-Scale Assessments		
Operational Feasibility & Logistics	Reviewing Data Collection Procedures	Tests the functionality of tablets, e-instruments and field data uploading processes to ensure smooth data capture.
	Enumerator & Test Administrator Training	Evaluates training effectiveness and identifies gaps in administrator preparedness.
	Understanding Student Behavior	Observes how students interact with assessment tools, including comprehension of instructions, response patterns and test timing and length.
Stakeholder Engagement & Policy Alignment	Alignment with Educational Goals	Ensures assessments align with national or international learning frameworks like SDGs, educational policies and relevant curriculum.
	Stakeholder Engagement	Engages policymakers, educators and communities to foster trust and contextual relevance.
Risk & Cost Management	Ensuring Reliability & Validity	Enhances statistical validation and construct refinement for reliable and valid assessments.
	Risk Mitigation	Identifies operational risks, biases and content misalignment before full-scale implementation.
	Cost Efficiency	Saves costs by preventing failures, minimizing revisions and optimizing data collection processes.
Capacity Building	Training & Readiness	Strengthens the skills of local teams through training sessions, feedback loops and hands-on experience, improving assessment execution.

Table 7 : Key Uses and Importance of Pilot Studies in Large-Scale Assessments

6.2 Steps in Conducting a Pilot Study for Large Scale Assessments

1. **Define Objectives:** Clearly articulate what the pilot aims to achieve, such as validating items, testing logistics or refining protocols.
2. **Design the Pilot:** Develop a smaller version of the full assessment with a representative sample of participants. Ensure alignment with the larger assessment framework.
3. **Select the Sample:** Choose a sample that mirrors the demographic and contextual diversity of the target population.
4. **Administer the Pilot:** Simulate the full assessment process, from test administration to data collection, monitoring and scoring.
5. **Analyze Data:** Use statistical methods to evaluate item performance, test effectiveness and data quality. Identify patterns and anomalies in student responses.
6. **Incorporate Feedback:** Gather input from test-takers, administrators and other stakeholders to refine the assessment process.
7. **Document Findings:** Prepare a detailed report outlining the successes, challenges and recommendations for the full-scale assessment.

6.3 Overall Considerations for Piloting FLN Assessment Tools

To ensure the effectiveness, validity and reliability of FLN assessment tools, various critical factors must be considered during the pilot phase. The table below provides a structured overview of these key considerations.

Key Considerations Regarding The Timing Of Pilots
• **School Calendar:** Schedule to avoid conflicts with school holidays, exams and events. • **Assessment Development Timeline:** Choose an optimal time of year, considering local factors such as climate, festivals, etc. Allow time for revisions between tool development and piloting, ensuring enough time for post-pilot analysis and tool refinement. • **Training of Assessors:** Conduct pilot after thorough training of assessors. • **Pre-pilot Preparations:** Allow enough time before the pilot to finalize logistics, school selection and material preparation. • **Sufficient Duration:** Ensure the pilot duration is sufficient to yield meaningful insights while also being short enough to minimize disruptions to the school schedule and learning process.
Key Considerations During Piloting
• **Assessment Content:** Ensure assessment content aligns with curricula, ensuring relevance and appropriateness. Ensure they are appropriate for the targeted grade level. • **Clarity of Instructions:** Ensure clarity of instructions for assessors and students (easy to understand, culturally and contextually appropriate). • **Representative Sample:** Select a representative student sample (age, gender, location, language). • **Scheduling and Logistics:** Coordinate logistics and ensure availability of materials. • **Consent and Permissions:** Obtain permissions from authorities and administrators. • **Testing Environment:** Set up a conducive, quiet and distraction-free testing environment. • **Training of Assessors:** Train assessors in standardized administration and scoring. • **Administration of the Assessment:** Follow standardized administration procedures, including the use of clear instructions, appropriate pacing and consistent scoring methods. • **Data Collection and Storage:** Use secure, reliable data collection tools. Ensure that the data is securely stored, backed up and protected to maintain confidentiality and integrity. • **Timing and Pacing:** Monitor timing and pacing to suit target age groups. • **Monitoring and Supervision:** Implement monitoring and supervision for quality control. Ensure that assessors are adhering to standardized procedures and that any issues or discrepancies are promptly addressed. • **Quality Control:** Regularly verify data accuracy and consistency. Address any data entry errors or inconsistencies immediately to maintain the quality of the data. • **Reliability and Validity:** Analyze reliability and validity across different assessors and settings. • **Feedback from Stakeholders:** Gather feedback from students, assessors and stakeholders for refinements and improvements.

Table 8 : Key Considerations while Planning a Pilot

6.4 Ensuring Assessment Effectiveness

Developing an assessment tool is a rigorous process that requires careful planning, iterative refinement, and thorough evaluation to ensure that test items are both valid and reliable. A well-constructed assessment tool accurately measures the intended constructs and produces consistent results across different contexts. For an assessment tool to be effective, it must demonstrate both validity and reliability. Pilot studies serve as an essential mechanism for evaluating these properties.

Assessment Effectiveness: Ensuring Validity, Reliability and Fairness

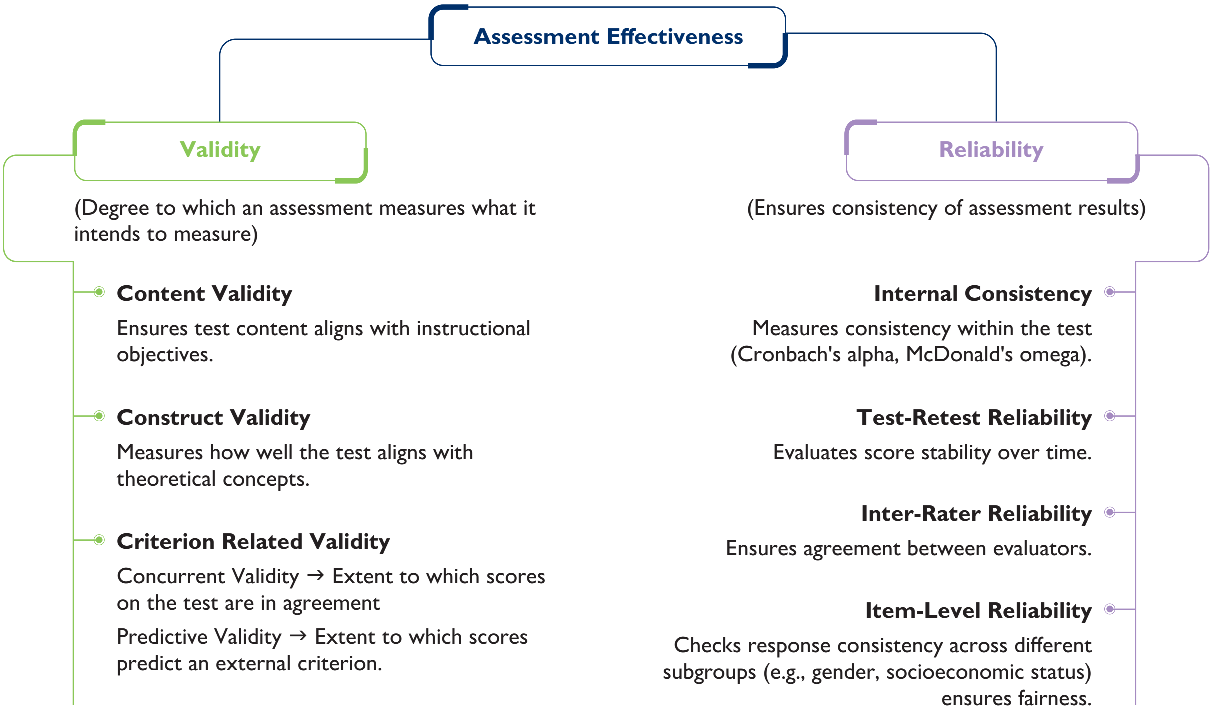


Figure 18: Types of Validity and Reliability Measures to Ensure Assessment Effectiveness

6.5 The Role of IRT in Analyzing Pilot Data

Item Response Theory (IRT) is a psychometric framework used to model the relationship between a person's latent trait (e.g., ability) and their probability of correctly answering an item. It is a model that describes how examinees respond to items on the assessment. Incorporating IRT in the analysis of pilot study data provides advanced psychometric insights, helping to refine test items and enhance overall assessment quality. **IRT plays a critical role in improving both validity and reliability by identifying weak items, assessing item difficulty and discrimination and ensuring fairness across diverse student populations.**

Some Advantages of IRT for Pilot Studies:

1. **Detailed Item Diagnostics:** IRT identifies item difficulty, discrimination and guessing parameters, enabling nuanced item evaluation.
2. **Key IRT models -**
    - **1PL Model (Rasch):** Focuses solely on item difficulty.
    - **2PL Model:** Adds item discrimination (how well an item differentiates between high- and low-ability individuals).
    - **3PL Model:** Includes a guessing parameter for multiple-choice items
3. **Ensuring Fairness through Person-Item Interaction:** IRT models individual ability estimates and their interaction with item parameters, ensuring that test items function fairly across different groups.
4. **Test Equating for Score Comparability:** IRT enables linking scores across different test forms, ensuring consistency in measurement over time and across various test versions, which is crucial for longitudinal studies.
5. **Scalability:** IRT results can be applied to various populations, ensuring cross-contextual validity.
6. **Reduced Biasness:** IRT reduces bias in the instrument through advanced techniques like differential item functioning.
7. **Vertical Scaling:** IRT can connect multiple levels of content, such as Math curriculum from Grades 3 to 12 if that is to be measured. This is known as vertical scaling within the progress monitoring framework.

6.6 Incorporating IRT in Pilot Data Analysis

6.6.1 Data Preparation

- **Scoring:**
    - Transform raw responses into formats suitable for IRT (e.g., dichotomous or polytomous).
- **Software:**
    - Use tools like R (e.g., ltm, TAM packages), Stata or commercial software like Winsteps or Mplus.
- **Sample size:**
    - Ensure adequate sample size for reliable parameter estimation (minimum 200–500 participants, depending on model complexity).

6.6.2 Steps in IRT Analysis

1. **Fit the model:**
    - Select an appropriate IRT model based on the assessment design.
    - Evaluate model fit using likelihood ratio tests or fit indices.

2. **Estimate Item Parameters:**
   - Calculate difficulty (b), discrimination (a), and guessing (c) parameters.
   - Identify poorly performing items for revision or removal.
3. **Assess Reliability:**
   - Use person separation indices (PSI) or test information functions to evaluate reliability. The Item Information Function shows the amount of information each item provides and it is calculated by multiplying the probability of endorsing a correct response multiplied by the probability of answering incorrectly.
4. **Detect Differential Item Functioning (DIF):**
   - Identify items that behave differently across subgroups to ensure fairness.

Item Information Function	$$\boxed{I_i(\theta, b_i) = P_i(\theta, b_i)Q_i(\theta, b_i)}$$	Where, θ = ability b_i = difficulty parameter Q = incorrect response b = correct response

5. **Generate ability estimates:**
   - Provide ability estimates (theta values) for each participant, independent of specific items administered.

6.6.3 Application of Results

- **Item Refinement:**
  - Modify or remove problematic items based on IRT diagnostics.
- **Test Length Optimization:**
  - Use IRT to identify the most informative items, potentially reducing test length without sacrificing reliability.
- **Standard Setting:**
  - Establish proficiency levels based on IRT-generated theta scores.

6.7 Data and Sample Requirements for Conducting Pilots for FLN Assessments

Minimum Data Requirement for Psychometric Analysis: At least 150 non-missing and nonzero scores are needed. These scores must fall within a reasonable and expected range similar to those in the final study.

Sampling Method: While random sampling is ideal, convenience sampling is typically used for pilots. It is ensured that selected students and schools should closely resemble the target population of the final assessment.

Challenges in Low-Performing Areas: In states/districts where 20%–30% of students score zero, a larger pilot sample is necessary to get enough nonzero scores. For example, if only 70% are expected to score above zero, a sample of around 200 students is needed to obtain 150 nonzero scores.

Oversampling Higher-Performing Schools: Oversampling ensures that sufficient nonzero data is collected for analysis.

Ensuring Diversity in the Pilot Sample:

- Include students from low, medium and high performing schools.
- If performance data is unavailable, use socioeconomic information to guide selection.

Pilot Sampling Size Per School:

- Unlike the final assessment (which limits to 10–12 students per grade per school), the pilot often samples more students per school.
- This allows faster data collection with fewer assessors.
- Acceptable practice as pilot data is not used for statewide performance estimates.

6.8 Analysis of Pilot Data for Establishing Reliability and Validity

Statistical analysis of pilot data for the FLN assessment tools is crucial for establishing their reliability and validity. It is essential that these analyses be conducted on both pilot data (for initial test operational data) and full study data (to determine whether or not any specific items should be removed from scoring).

Here are some key statistical analyses that should be conducted:

- **Descriptive Statistics:** Calculate basic descriptive statistics such as mean, median, standard deviation and range for each assessment item and overall scores. This helps to understand the central tendency and dispersion of the data.
- **Item Difficulty:** Analyze the difficulty level of each assessment item by calculating the proportion of students answering each item correctly. Items that are too easy or too difficult may need to be revised or replaced.
- **Item Discrimination:** Calculate the item discrimination index, which shows how well an item can differentiate between high-performing and low-performing students. Items with low discrimination indices may need to be revised or replaced.
- **Internal Consistency:** Measure the internal consistency of the assessment using reliability coefficients such as Cronbach's alpha. This helps to determine how consistently the items within each domain of the Literacy and Numeracy tools are measuring the same underlying construct.
- **Interrater Reliability:** If multiple assessors are involved, calculate the interrater reliability to determine the degree of agreement between assessors in scoring the students. High interrater reliability indicates that the scoring guidelines are clear and consistently applied by different assessors.
- **Factor Analysis:** Conduct factor analysis, such as exploratory factor analysis (EFA) or confirmatory factor analysis (CFA), to examine the underlying structure of the assessment items and ensure that they are measuring the intended constructs.
- **Construct Validity:** To ascertain construct validity, item-level statistics should be produced to

ensure that all items are performing as expected. Rasch analyses (which rely on an assumption of unidimensionality) provide construct validity information in several ways. First, the Rasch model places items and students on the same scale of measurement, in order, from easy (low ability) to difficult (high ability). Therefore, the order of the items from least to most difficult is the operational definition of the construct. If this definition matches the intended design, there is an indication of construct validity. However, if there are instances where students do not have representative items accurately assessing their ability, it is said that there is an underrepresentation of the construct.

- **Item Fit:** Rasch analyses assess item performance through fit statistics. If the items are not accurately measuring ability, or are producing "noise," then they will have higher statistics (≥ 2.0) indicating misfit and will need to be reevaluated. Assessments with many misfitting items are said to have constructed irrelevant variance, which is also a detriment to the construct validity.

- **Differential Item Functioning (DIF) analysis:** Conduct DIF analysis to identify any potential bias in the assessment items across different subgroups of students (e.g., gender, language,

Statistical Indices Used for Establishing the Reliability and Validity of Tools & Threshold Values

1. **Test Reliability:** Reliability is defined as the overall consistency of a measure. The most widely used measure of test-score reliability is Cronbach's alpha, which is a measure of the internal consistency of a test. For Cronbach's alpha or other measures of reliability, the higher the alpha coefficient or the simple correlation, the less susceptible the tool scores are to random daily changes in the condition of the test takers or of the testing environment. As such, a value of 0.7 or greater is seen as acceptable.

2. **Item Difficulty (as per Rasch Model)**: Difficulty estimates range from -3 to +3 from most easy to most difficult. It is in line with the international best practice that there should be enough difficult items at the top end as well as enough easy items at the bottom end of the scale to target the students across the entire range of ability distribution and provide reliable information on the measured constructs.

3. **Infit Statistics:** If the items are not accurately measuring ability, or are producing "noise," then they will have higher statistics (≥ 2.0) indicating misfit and will need to be reevaluated. Assessments with many misfitting items are said to have irrelevant construct variance, which is also a detriment to the construct validity. It is international best practice to use a weighted mean-square fit statistic and a reasonable range of 0.8-1.2, which indicates that an item does not deviate significantly from the model.

4. **Point Biserial correlation (rpb.WLE):** This is an indicator of correlation of an item score with test score and therefore serves as an indicator for item discrimination and reliability. As an international best practice, items with WLE lower than 0.2 are flagged.

6.9 Challenges and Limitations of Pilot Studies

While pilot studies provide essential insights into assessment effectiveness, they also come with several challenges:

- **Resource Constraints:** Conducting pilots requires substantial time, funding and trained personnel, which may strain available resources.

- **Sampling Bias:** If the pilot sample is not representative of the target population, results may not generalize effectively.

- **Over-interpretation of Results:** Treating pilot outcomes as definitive rather than exploratory can lead to flawed decision-making.

- **Limited Scope:** Pilots may not fully capture challenges that arise at scale, potentially overlooking key implementation issues.

Additionally, the use of Item Response Theory (IRT) in pilot studies presents specific challenges:

- **Technical Expertise:** Implementing IRT requires specialized knowledge in psychometrics and statistical modeling.
- **Computational Demands:** Estimating IRT parameters can be resource-intensive, particularly for large-scale assessments.
- **Sample Size Considerations:** Small or sparse data sets can result in unstable parameter estimates, reducing reliability.
- **Multidimensionality Complexity:** Assessments that measure multiple constructs require more sophisticated IRT models, increasing analysis complexity.

Pilot studies and field trials are indispensable in the planning and implementation of large-scale assessments. They provide a structured approach to identifying and addressing potential challenges, ensuring that the assessment tools and processes are robust, efficient and equitable. By investing in rigorous pilot studies, stakeholders can enhance the credibility and impact of large-scale assessments, ultimately contributing to improved educational outcomes and evidence-based policymaking.

7 SAMPLING

7

Sampling

7.1 Econometric Approach

7.1.1 Requirement of Sampling

Large-scale assessments are 'sample-based', wherein a select few students from each block or district are administered the assessment. Since these students are representative of the entire block or district, it is important to get a well-representative sample. Sampling is the statistical process of selecting a limited number of observations from a large population. The study results obtained from a sample may be generalized to the wider population when certain statistical considerations are met.

7.1.2 Considerations for Sampling

The sample size required to ensure that it is adequately representative and an accurate estimation of the underlying population being studied is determined by several statistical factors as below.

Statistical Significance: This is the probability at which the results and inferences from a sample can be applied to the population. The commonly accepted levels of statistical significance in quantitative research are 99%, 95% and 90%.

Statistical Power: Power defines the probability that the "no-effect" observed on the sample can be generalized to the population. In the case of periodic state-level or district-level FLN surveys, no effect would imply determining that the mean learning levels are not statistically different for the same geography at different points in time. Power is usually set to 80% in educational surveys.

Minimum Detectable Effect (MDE): MDE refers to the size of the change or difference we can detect using our sample. MDE may refer to the difference between sample statistics and population parameters, the difference between any two groups or it could be the difference between the same group over time (e.g., between baseline and endline). For example, a 5%

MDES means that the smallest effect that can be measured through the sample is 5% - effects greater than 5% can also be captured through this sample. It also directs us to the point that larger samples are required to detect lesser effect sizes. Stakeholders are informed about the MDE beforehand.

Design Effect: When cluster sampling is used, a multiplicative correction factor known as the design effect is applied to the sample size calculation. Cluster sampling is often used in surveys to improve logistics and implementation, particularly in educational research where students, typically grouped in clusters such as schools, are the primary sampling units. These clusters often share similar background characteristics, leading to a positive correlation among observations within the same cluster. This correlation, known as the Intra-Class Correlation (ICC), reflects the similarity of individuals within clusters. The design effect accounts for this correlation and adjusts the sample size accordingly. The design effect is calculated using the ICC and the average cluster size and it is incorporated into the sampling formula. It adjusts the required sample size for clustered sampling compared to simple random sampling. For example, if the design effect is 2, the required sample size is doubled compared to what would be needed under simple random sampling.

The design effect can be quite high if there are multiple layers of clustering. In complex survey designs where multiple levels of clustering are involved — such as students within classes, classes within schools, schools within blocks, blocks within districts and districts within states — the design effect can increase significantly. For instance, the National Achievement Survey (NAS) 2021 utilized a design effect of 7.

Hypothesis Testing: Hypothesis testing is a statistical method used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population. In this framework, the null hypothesis (H_0) represents the default assumption — typically, that there is no effect or no difference — while the alternative hypothesis (H_1) proposes that there is an effect or a difference. Researchers collect data and calculate a test statistic, which is then compared to a critical value or used to derive a p-value. If the p-value falls below a predetermined significance level (such as 0.05), the null hypothesis is rejected in favor of the alternative, suggesting that the observed effect is statistically significant.

7.1.2.1 Understanding Statistical Errors: False Positives and False Negatives

An alternative way to grasp the concepts of statistical significance and power is by considering false positives and false negatives:

- **False Positives (Type I Error):** A false positive occurs when the test incorrectly detects an effect that does not exist, rejecting a true null hypothesis. The significance level (alpha) is the probability of making this error. For example, if alpha is set at 5%, there is a 5% chance of concluding that an effect exists when it does not.

- **False Negatives (Type II Error):** A false negative happens when the test fails to detect a real effect—failing to reject a false null hypothesis. Statistical power, defined as 1 minus the probability of a Type II error (beta), represents the likelihood of correctly identifying a true effect. A power of 80% means there is a 20% chance of a false negative or conversely, that the test will correctly detect a true effect 80% of the time.

In the context of sampling, ensuring a sufficiently large sample size is key to minimizing false negatives by increasing power. Simultaneously, setting a stringent significance level helps control the rate of false positives. Balancing these errors is critical: a sample that is too small may fail to detect important differences (increasing false negatives), while a lenient significance threshold might lead to detecting effects that are not truly present (increasing false positives). Together, these concepts underscore the importance of robust sample size calculations and thoughtful survey design.

7.1.3 Sample Size Calculation

The sample size should be decided based on the purpose of the study. State FLN assessments fall into the category of periodic studies for informing the decision-makers about the prevailing situation and changes over time. These studies can further inform intervention planning and diagnostics. Given the purpose of this research, a non-experimental research design without a counterfactual can be followed. The sample size calculation provides the overall number of students required for the survey. The requisite number of sample schools can be obtained by dividing the overall sample size with the size of each cluster, viz., the number of students required in each school.

For the ease of consolidation of the sampling across different geographical markers and assumptions, the following table should always be maintained for each sampling exercise.

States	No. of Districts	Standard Sampling Assumptions	District Representative Sample

Table 9 : Sampling Table

Calculation of Sample Size

MDE – Based Formula

$$n = \frac{D\,[Z_{1-\alpha}\sqrt{20P(1-P)} + Z_{1-\beta}\sqrt{P_1(1-P_1)+P_2(1-P_2)}]^2}{(P_2-P_1)^2}$$

P_1 = Estimated proportion at baseline
P_2 = Proportion expected at the time of the survey
P = Average of P_1 and P_2
$MDE = P_2 - P_1$
$Z_{1-\alpha}$ = z-score corresponding to the level of significance
$Z_{1-\beta}$ = z-score corresponding to statistical power
D = Design effect

Used when the objective is to detect a specific minimum difference (or effect size) between two groups or over time, making it more suitable fo comparative or experimental study designs. Main educational research studies - including the recommended FLN surveys use this formula.

Cochran's Sample Size Formula

$$n_0 = (Z^2 \times p \times (1 - p)) / e^2$$

where:

- **n_0: Estimated sample size**
- **Z: z-score corresponding to the desired confidence level (e.g., 1.96 for 95% confidence)**
- **p: Estimated proportion of the attribute in the population (if unknown, use 0.5 to maximize sample size)**
- **e: Desired margin of error (e.g., 0.05 for 5%)**

Used when the goal is to estimate a single proportion with a predetermined level of confidence and precision, particularly in large or effectively infinite populations. It is ideal for cross-sectional surveys where the focus is on measuring the prevalence of an attribute rather than detecting changes between groups.

Figure 19: Sample Size Calculation Using MDE Based Formula and Cochran's Formula

MDE Based Sampling Formula Application:	Cochran's Sample Size Formula Application:
Let's assume the following values: • P_1 (Baseline proportion of students achieving grade-level proficiency): 0.30 (30%) • P_2 (Expected proportion after intervention): 0.40 (40%) • $P = (P_1 + P_2) / 2 = (0.30 + 0.40) / 2 = 0.35$ • $MDE = P_2 – P_1 = 0.10$ (10 percentage points) • $Z_1\ \alpha$ (for 95% confidence level): 1.96 • $Z_1\ \beta$ (for 80% power): 0.84 • D (Design effect): 2 Applying to the formula: $n = D \times \left[Z_{1-\alpha}\sqrt{2P(1-P)} + Z_{1-\beta}\sqrt{P_1(1-P_1)+P_2(1-P_2)}\right]^2 /(P_2-P_1)^2$ • Compute the first term + second term within brackets and adding them: $= (1.321+0.564)_2=(1.885)_2 \approx 3.553$ • Divide by $(P_2 – P_1)^2$: $(P2–P1)_2=(0.10)_2=0.01$ • Apply the Design Effect: $n = 2\times(3.553/0.01) = 2\times355.3 = 710.6 \approx 711$	For example, at a 95% confidence level ($Z = 1.96$), with $p = 0.5$ and $e = 0.05$: $n_0 = (1.96^2 \times 0.5 \times 0.5) / (0.05^2)$ ≈ 384 For a finite population, the finite population correction (FPC) is applied: $n = n_0 / [1 + ((n_0 – 1) / N)]$ where N is the total population size. If $N = 2000$: $n \approx 384 / [1 + ((384 – 1) / 2000)]$ ≈ 322

7.2 Sample Selection

7.2.1 Selection of Level of Representativeness

The key objective of statistical sampling is to ensure that results obtained from the sample may be generalized to the wider population. Keeping this in mind, it is important to identify the level of representativeness required from the sample. For instance, if the study requires a district-level estimation or comparison in learning scores, the sample size is computed for the district.

The remainder of this section discusses the selection of schools and students with the assumption that the purpose of the educational survey is to provide district-level estimates.

7.2.2 Selection of Schools

Once the level of representativeness of the sample is determined - which, for this discussion, is assumed to be the district - the next step is to select schools within the district for undertaking the surveys. Given that their selection forms the first step in the process of sample selection, schools constitute the ***primary sampling unit (PSU)*** in this case.

As discussed earlier, the sample size calculations will specify the overall number of schools and the count of students per school to be selected for the survey.

The complete set of schools within the district from which a sample is to be selected forms the ***sampling frame***. The U-DISE+ dataset compiled by the state each year would provide updated information on this account - it can, therefore, be used as the sampling frame for school selection. Schools are then chosen from the sampling frame through a process of ***random selection.***

One key factor to keep in mind is that a sizable proportion of schools in the underlying population may fall into certain differential criteria - such as type of schools (primary only, primary and upper primary and so on) or school management (including government-run and government-aided schools). It may be important to integrate these criteria into the selection of schools so that the resultant sample is adequately representative of the population. This is ensured through a technique called ***stratified sampling.***

Stratified sampling divides the overall population into sub-groups called strata and randomly selects sampling units from each sub-group. For the FLN surveys, key criteria for the creation of strata may include the following:

1. **Type of School:** This can cover types such as primary schools and primary with upper primary schools
2. **School Management:** Government-run, government-aided and private schools, among other categories
3. **Language of Instruction:** It would be useful to further divide schools based on the instruction medium, particularly to take into account the use of regional/tribal languages
4. **Schools Catering to Particular Social Groups:** In particular, schools with a high proportion of students from Scheduled Tribes could be considered as a sub-group

Stratified sampling requires the availability of information on the prevalence of characteristics used to determine the strata. This information is used to determine the proportion of schools that must be chosen from each stratum out of the overall sample size.

For instance, let us suppose that a district-level FLN survey requires the selection of 100 schools. Within the district, 76% of primary grade students attend primary-only schools, while the remaining 34% attend primary-with-upper-primary schools. Stratified school selection would require randomly sampling 76 schools from the primary-only category and 34 primary-with-upper-primary schools.

Unforeseen circumstances at the time of the surveys, such as the absence of teachers, mandate that school selection takes into account a certain number of buffer/replacement schools (usually 5% to 10%). The selection of replacement schools should also be carried out randomly.

7.2.3 Sampling Units

Student selection forms the first step of the sample selection process - hence, students comprise the primary sampling units (PSUs) in the FLN surveys. The selection of students for FLN surveys within the sampled schools is carried out by the Field Investigators (FIs) and supervisors on school premises at the time of data collection. Student selection should be carried out randomly.

The following key points may be kept in mind so that student selection is safeguarded against the possibility of selection bias:

- The sample of students within a school should be chosen by the FIs/data collectors/enumerators and should not be influenced by the school officials
- Class lists/attendance registers maintained by the teachers may be used for randomly selecting the students
- In case these are unavailable, selection should be carried out in such a manner that bias is minimized. For instance, FIs should ensure that students seated in the front row of the classroom are not the only ones selected. Similarly, students suggested by teachers should not be selected
- To the extent possible, an equal proportion of boys and girls should be selected

7.2.4 Selection of Students

Within each selected school, students are the primary sampling units (PSUs):

- **Random Sampling:** Field investigators (FIs) should randomly select students from the relevant grade (using class lists or attendance registers where available). Wherever such lists are absent, a stratified random sampling should be followed using roll numbers or student seating patterns in the classroom.
- **Avoiding Bias:** Ensure that selections are not influenced by school officials or limited to convenient groups (such as students in the front rows).
- **Ensuring Balance:** Aim for an equal representation of boys and girls.
- **Weighting:** Assign sampling weights to students based on the inverse of their selection probability to adjust for any differences in representation.

7.3 Step-by-Step Sampling Guide

Adapted from the step-by-step sampling guide, the following detailed steps outline the entire sampling process for district-level assessments.

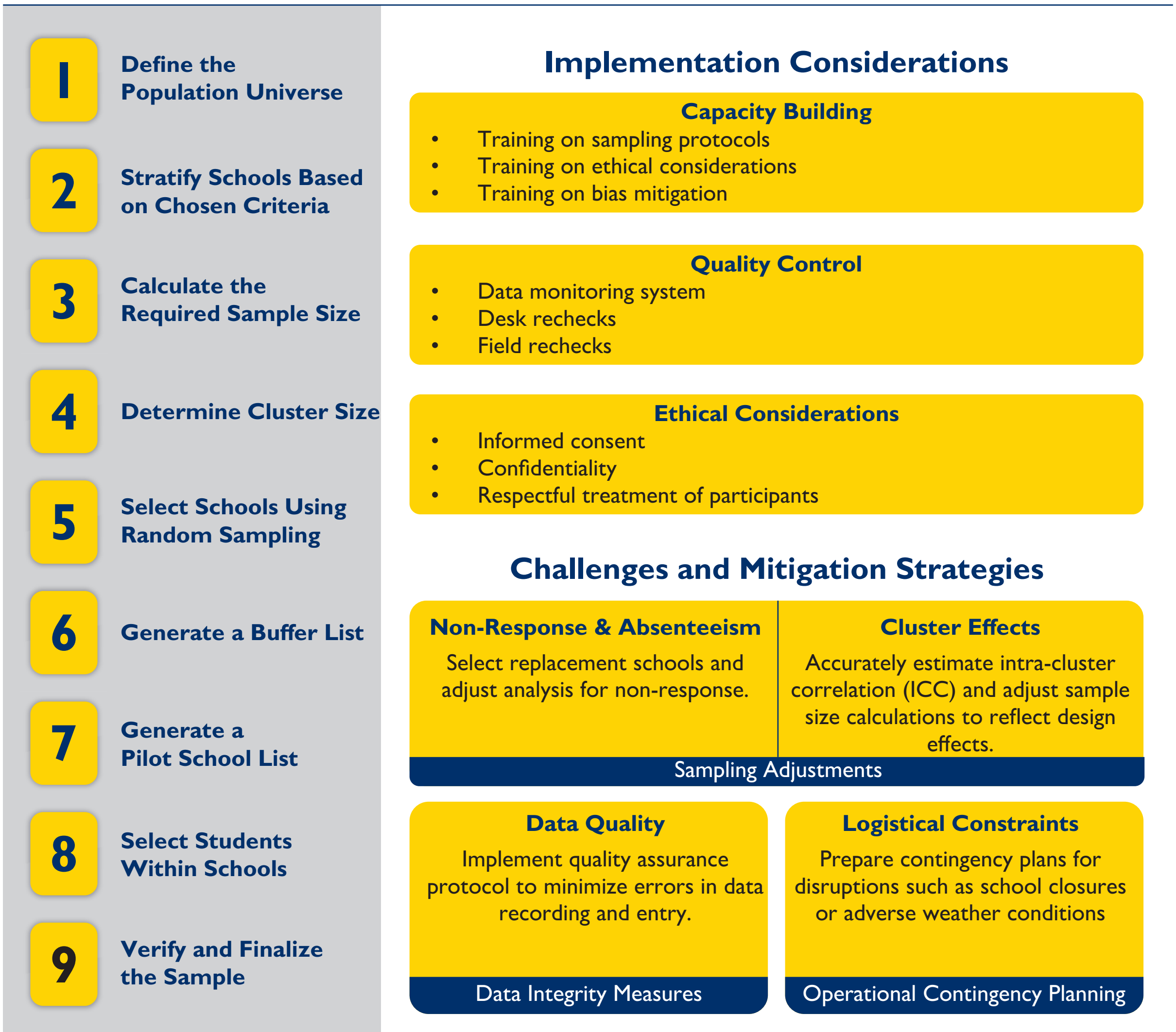


Figure 20: Step-by-Step Sampling Guide

7.3.1 Define the Population Universe

- **Data Source:** Use the latest UDISE data to obtain a comprehensive list of schools in the district.
- **Cohort Identification:** Identify the current cohort for the grade of interest (e.g., if the data is one year old, use Grade X–1 enrollments as a proxy for the current Grade X).
- **Preliminary Filtering:** Decide on stratification criteria and remove schools lacking complete data (e.g., missing Rural/Urban or Management Type information). Apply any additional state-approved filters.

7.3.2 Stratify Schools Based on Chosen Criteria

- **Criteria Selection:** Stratify the schools using factors such as Rural/Urban status, management type and school size.
- **Adjusting Strata:** If any stratum has too few schools, consider combining it with a related stratum to ensure that each group has at least 2–3 schools. This step maintains the proportional distribution of schools in the district.

7.3.3 Calculate the Required Sample Size

- **Statistical Parameters:** Use a 95% confidence level and a 5% margin of error for the initial calculation.
- **Finite Population Correction (FPC):** Adjust the calculated sample size based on the total number of schools in the district.
- **Proportional Allocation:** Allocate the sample across strata to reflect their proportional representation in the district.

7.3.4 Determine Cluster Size

- **Enrollment Data Extraction:** Retrieve grade-wise enrollment data from UDISE for each school in the sample.
- **Adjustment for Absenteeism:** Modify enrollment figures using district or state-level absenteeism rates to estimate the number of students likely to be present on assessment day.
- **Potential Sample Calculation:** Identify the maximum number of students that can be feasibly assessed per school.
- **Realistic Cluster Size:** Compute the average potential sample size across schools and validate this figure with the state and assessment teams to ensure operational feasibility.

7.3.5 Select Schools Using Random Sampling

- **Main School List:** Randomly select schools based on the calculated sample size and stratification criteria.
- **School Weights:** Generate the main list and compute school weights based on the probability of selection.

7.3.6 Generate a Buffer List

- **Backup Selection:** From the remaining schools, randomly select a buffer list (typically 5–10% of the sample) to serve as replacements if needed.
- **Weighting:** Calculate weights for the buffer list similarly to the main list.

7.3.7 Generate a Pilot School List

- **Pilot Objectives:** Select pilot study areas through state discussions, aiming to cover a range of student proficiency levels.
- **Random Selection within Areas:** Randomly select schools within these designated pilot regions.
- **Sample Size for Pilot:** Target at least 150 responses per competency (oversampling to 200–230 students per grade) and typically select 25–30 schools.
- **Exclusivity:** Ensure that pilot schools do not overlap with the main or buffer lists to maintain study integrity.

7.3.8 Select Students within Schools

- **Random Student Selection:** For each selected school, randomly sample students from the grade of interest.
- **Weight Assignment:** Compute student-level weights using the inverse probability of selection. This often involves a two-stage process—first for school selection (weighted by enrollment category) and then for individual student selection.

7.3.9 Verify and Finalize the Sample

- **Representativeness Check:** Ensure that the final sample mirrors the district's school population distribution.
- **Random Checks:** Conduct verification to confirm that no particular school type or location is over- or under-represented.
- **Finalization:** Lock the main, buffer and pilot school lists and share them with the state team.

Note: Limiting the number of strata to 2–3 is essential; too many strata can force the selection of a minimum number of schools per stratum, artificially inflating the overall sample size.

A robust sampling strategy is vital for producing reliable, generalizable findings in educational assessments. By integrating econometric principles, applying rigorous statistical methods (including Cochran's formula) and following a detailed, step-by-step procedure, researchers can design surveys that accurately reflect the target population. Attention to implementation and proactive challenge management further enhances the quality and credibility of the survey outcomes, providing a solid empirical foundation for policy-making and educational interventions.

Case study : State-Level Sampling Strategies and Operational Challenges in Educational Assessments: Four Case Examples from 2025

Case Example 1 – Block-Level Coverage with District-Level Limitations

In one state, the 2025 assessment adopted a block-level sampling strategy in which all blocks within each district were included, with 10 schools selected per block. This approach was chosen to align with the state's existing planning and reporting structures, which are organized at the block level. Sampling only a subset of blocks would have been poorly received by local stakeholders and would have created reporting gaps. However, a notable challenge emerged in districts with fewer blocks, as the total number of sampled schools per district was relatively low. This resulted in insufficient statistical power for generating precise district-level estimates. While the method ensured full geographic and administrative coverage, it compromised the accuracy of findings in some districts due to limited sample sizes.

Case Example 2 – Language Complexity and Operational Simplification

In another state, the target population was initially defined based on the medium of instruction, leading to the identification of 45 unique language combinations across schools. These included multiple dominant and indigenous languages, often used in combination as mediums of instruction. The complexity of this multilingual landscape posed operational challenges, particularly in identifying field investigators proficient in all required language combinations. To streamline the process, the sampling design was simplified by fixing the first language (L1) at the district level: any medium of instruction used in over 25% of schools within a district was considered L1. If two mediums met this criterion, both were treated as L1. Further refinement involved excluding very small schools, based on minimum enrollment thresholds—at least 4 students for the more commonly used mediums and at least 2 for less prevalent ones. Since schools below these thresholds represented only a small share of the overall population, their exclusion improved operational feasibility without significantly compromising representativeness. Additionally, access challenges in certain districts limited coverage to only a subset of blocks, further narrowing the effective sample in those regions.

Case Example 3 – Balancing Medium of Instruction and Urban-Rural Representation

A third state had to account for two main mediums of instruction, with a third language taught as a second language. While one medium was widely distributed across all districts, the other was limited in presence and concentrated in select areas. To maintain balance between representation and resource constraints, the sampling design capped the number of schools per district at 50 (with one urban district as an exception). These schools were proportionally distributed across four strata: rural and urban schools for each of the two primary instructional languages. In the major urban district where both language groups had a strong presence, a larger sample of 92 schools was drawn, equally divided between the two language-urban strata. Due to the sparse distribution of schools offering the less prevalent medium, district-level estimates for this group were not feasible and data was reported only at the state level. This design ensured state-level insights while acknowledging district-level limitations.

Case Example 4 – Targeted Sampling Based on School Management

In the fourth case, the sampling population was defined based on school management type, in line with the state's educational priorities and operational capacity. Out of 14 possible management categories, three were selected for inclusion:

- Government schools under the primary education department
- Government-aided schools
- Schools managed by departments serving scheduled tribe and scheduled caste communities

This focused scope helped align the assessment with the state's strategic goals while ensuring operational feasibility and a manageable sample size.

8 DATA COLLECTION

8

Data Collection

Data collection is a crucial phase in FLN assessments, requiring thorough planning and adherence to best practices to ensure high-quality outcomes. The data collection plan includes three key components- Planning for logistics, selecting and training field investigators (FIs) and managing field operations.

a. Planning for logistics involves preparing for effective data collection, including estimating the number of field investigators required, determining the number of days needed for fieldwork, reviewing budget requirements and planning around these operational details to ensure a smooth process. This detailed logistical planning lays a strong foundation for the next components — selecting and training field investigators and managing field operations, to ensure accuracy, consistency and efficiency throughout the data collection phase.

b. Selecting and training FIs is essential to ensure accuracy and consistency in data gathering. During fieldwork, it is essential to follow the standardized protocols, prioritize children's safety and well-being, protect the integrity of assessment tools and manage data responsibly with a focus on the quality, confidentiality and fidelity of the data collected.

c. Finally, managing field operations covers scheduling and coordinating activities and defining roles and responsibilities to ensure smooth execution. This phase also includes regular monitoring of the data to promptly identify any discrepancies or instances of non-adherence to protocols, allowing for corrective actions to maintain data quality and consistency throughout the assessment process.

8.1 Planning for Logistics

Good planning for data collection is crucial as poor preparation can lead to several negative consequences, including budget overruns and low-quality data, both of which can compromise the success of the entire project. Effective logistics planning involves multiple components.

Estimating the Number of Field Investigators

The number of field investigators required is determined by the number of schools to be assessed in each district. States with fewer districts will require fewer FIs, whereas states with more districts will necessitate a larger team of FIs to cover the workload effectively. Additionally, estimating the number of FIs required is also dependent on how many days the assessment will go on for, whether the FIs can move across districts, how many students can be assessed in a day (which depends on the length of the tool, number of subjects to be assessed, the length of the school day, etc.). An example is discussed in Annexure 3.

Good logistical planning ensures a balance between these factors, enabling high-quality data collection while remaining within budget.

Additional Planning Steps for Smooth Execution

Once the number of field investigators (FIs) has been determined, several additional planning steps are essential to ensure the smooth execution of the assessment. These include:

1. **Setting Targets**
   - **Student sample size:** Define the target number of students to be assessed per school, ensuring that the sample size is representative and aligns with the study's overall objectives.

2. **Budget planning**
   - **Remuneration:** Decide on the payment structure for FIs, whether based on the number of days worked or the number of schools visited.
   - **Trainer logistics:** Allocate a budget for trainers, covering their travel, accommodation and food.
   - **Daily allowances:** Include provisions for meal allowances for FIs during their field visits.
   - **Material costs:** Budget for printing assessment materials and other necessary documents, such as survey forms or instructional guidelines.

3. **Verification and logistics**
   - **Verify sample schools:** Confirm the selected sample schools and identify their exact locations using the help from state resource persons, ensuring accuracy in the sampling process.
   - **Authorization from state :** Ensure a written permission from the state department for data collection.
   - **Travel arrangements:** Coordinate transportation and accommodation for FIs, addressing any logistical requirements to avoid delays.
   - **Replacement protocols:** Develop clear criteria for replacing schools, teachers or students in cases of unavailability or unforeseen issues, ensuring the sample remains valid.

4. **Itinerary development**
   - **Detailed schedules:** Prepare a comprehensive schedule that includes specific dates, the list of schools and head teacher contact information.
   - **Local context consideration:** Involve someone familiar with the region in itinerary development to ensure feasibility, taking into account local festivals, holidays and weather conditions to avoid overlaps or disruptions.

8.2 Selection and Training of Field Investigators

Field investigators are personnel who will be responsible for collecting data directly from students, schools and teachers during the assessment process. Training of field investigators is a crucial step in ensuring the successful implementation of FLN assessments. The objective of the training is to familiarize the investigators with standardized protocols, assessment tools and instructions and to equip them with the necessary skills to administer the assessment effectively and consistently across different regions.

FI Selection for Training :

FIs are selected based on their performance in a preliminary assessment, to evaluate their suitability for the role. They should have the following minimum qualifications and skills :

- Fluently read and speak the language being assessed and the language in which instructions are provided to pupils.
- Have experience working with young children.
- Be able to follow directions and pay attention to detail.
- Have experience using a smartphone.
- Have previous experience participating in data collection.
- Work well as part of a team. Additionally, they should demonstrate they can work well in a school environment, which requires following certain protocols, respecting school personnel and property and interacting appropriately with students.
- Be available for the entire duration of data collection.
- Be able to travel to rural and remote areas.
- Be employed by a relevant educational institution (if applicable).
- Demonstrate ability to accurately administer the assessment during the training.

8.2.1 FI Training Plan

The training plan for field investigators must be comprehensive and structured to ensure effective preparation for the FLN assessments. It should include the following key components:

1. **Training Model:** A cascaded training model is employed to scale training efforts across large regions. In this approach, Trainers (MTs) Trainers first undergo intensive training and then take on the responsibility of training Field Investigators (FIs) in various districts. This hierarchical structure enables the efficient and widespread dissemination of knowledge while ensuring the consistent application of standardized protocols and assessment instructions.
2. **Training Locations:** Identify the number of locations where the training will take place, considering logistical convenience and accessibility for all participants.
3. **Number of Participants:** Specify the number of field investigators to be trained at each location to ensure adequate resources and manageable group sizes.
4. **Number of Master Trainers:** For each training venue, it is recommended to have at least three Master Trainers (MTs) per cohort of 20 FIs to ensure sufficient supervision and support. This ratio allows for effective guidance and ensures that all FIs receive personalized attention. Additionally, it is important to include at least one female MT in each cohort to foster a comfortable environment for female participants and facilitate discussions on sensitive topics, such as the Prevention of Sexual Harassment (POSH).

5. **Duration and Agenda:** Outline the number of days required for the training and provide a detailed agenda for each day. The training for FIs should ideally span over two to three days, focusing on administering tasks for all subjects and mock testing with an inter-rater reliability exercise. Depending on the number of grades being assessed, the training will include specific sessions on grade-specific toolkits. For instance, if multiple grades are being assessed, additional time will be incorporated into the training plan to ensure that FIs are familiar with the relevant toolkit for each grade. This will be reflected in the adjusted duration of the training, ensuring that each grade's specific tools are adequately covered. Thus, the training structure is designed to ensure that all participants are well-prepared to conduct FLN assessments effectively. The agenda should cover:
    - Overview of FLN assessments and objectives
    - Introduction to standardized protocols and assessment instructions
    - Practical training on using assessment tools
    - Role-playing and mock assessments to reinforce consistency in data collection
    - Sessions on managing data quality and ensuring adherence to protocols
    - Q&A sessions for clarifications and troubleshooting
6. **Budget:** Include a detailed budget covering training expenses such as venue rental, catering and refreshments, rental of AV devices such as projectors/LCD screen, lodging and transportation of MTs, print materials and participant allowance. Small contingency fund should be included to cover any unforeseen expenses.
7. **Training Materials and Resources:** List the materials needed for the training, including:
    - Assessment toolkit including stimulus sheets and worksheets
    - Contextual tools
    - Instructions for task administration
    - App for data collection
    - Training PPT
    - List of Frequently Asked Questions (FAQs) to assist with common issues
    - Guidelines for child safety and ethical considerations
    - Audio-visual resources for presentations
    - Handouts or reference materials for participants

8.2.2 FI Training Workshop Agenda

The objectives of the training workshop are:

a. To emphasize the importance of collecting reliable data in the survey, ensuring that FIs understand their role in capturing accurate data that reflects the true learning levels in the state.

b. To clarify the roles and responsibilities of FIs, covering all standard operational protocols and procedures to be followed before, during and after the assessment.

c. To explain the random sampling method for selecting students within each school and grade for the assessment, ensuring that FIs understand how to apply this process accurately to uphold the integrity and representativeness of the sample.

d. To provide an overview of the assessment tools as well as the background questionnaires for pupil, teachers and schools ensuring familiarity with the instruments.

e. To train FIs on administering the assessment tools, focusing on standardized procedures and protocols for conducting assessments and recording data.

f. To introduce the data collection app, ensuring that FIs understand how to record, save and upload data correctly using the application.

g. To conduct live demonstrations of assessment administration and providing FIs with opportunities for hands-on practice followed by feedback and guidance to ensure correct implementation.

h. To conduct mock testing for FIs to assess their preparedness post-training, establish inter-rater reliability among Fis and provide feedback on any discrepancies observed during the mock assessments. This will help refine their understanding of the process and improve consistency in data collection.

i. To select the final FIs based on their performance in the quiz, mock tests, practice tasks and the inter-rater reliability exercise, ensuring that only the most qualified FIs are chosen for the fieldwork.

j. To identify the best-performing FIs to be designated as field supervisors, based on their demonstrated leadership, communication and data collection skills, ensuring effective supervision and quality control in the field. Field supervisors may either be selected from among the FIs or if a cadre of cluster/block/district coordinators employed by the state is available, they can also be utilized for this role. However, it is recommended that these coordinators attend the entire training to effectively support the FIs in the field.

k. To establish clear criteria for assigning schools to FIs and organize the allocation process. Define the typical daily routine for FIs in schools, including the sequence of activities they will follow during fieldwork.

एजेंडा - Day 1			
9:00 AM	10:00 AM	1	विद्यार्थी लर्निंग आउटकम सर्वेक्षण के मुख्य बिन्दु : क्यों आवश्यक है , कब किया जाएगा, कहाँ किया जाएगा, किनका किया जाएगा, कैसे किया जाएगा FI के रूप में आपके कर्तव्य - डेटा की सटीकता और विश्वसनीयता
10:00 AM	11:00 AM	2	स्कूल में जाकर क्या करें • स्कूल में बातचीत • बच्चों का चुनाव - चुनाव की प्रक्रिया • तैय्यारी - उपयुक्त स्थान का चुनाव • ध्यान में रखने वाली महत्वपूर्ण बातें
11:00 AM	11:30 AM	3	इस आकलन में क्या क्या जांचा जाएगा
11:30 AM	12:30 PM	4	आकलन कैसे करें • सामग्री की समझ • डाटा कलेक्शन app की समझ - app इंस्टालेशन and रजिस्ट्रेशन
12:30 PM	1:30 PM	5	भाषा के आकलन की समझ - प्रत्येक प्रश्न को कैसे पूछें और उत्तर कैसे app में दाखिल करें - कक्षा 2
Lunch Break			
2:30 PM	4:30 PM	6	भाषा के आकलन की समझ - प्रत्येक प्रश्न को कैसे पूछें और उत्तर कैसे app में दाखिल करें - कक्षा 3,4, डाटा को सेव कैसे करें और कैसे हम तक पहुंचाएं
4:30 PM	5:30 PM	7	Queries and revision

एजेंडा - Day 2			
9:00 AM	10:00 AM	1	दोहराव
10:00 AM	11:00 AM	2	गणित के आकलन की समझ - प्रत्येक प्रश्न को कैसे पूछें और उत्तर कैसे app में दाखिल करें डाटा को सेव कैसे करें और कैसे हम तक पहुंचाएं
11:00 AM	1:00 PM	3	बैकग्राउंड प्रश्नोत्तरी कैसे भरें • बच्चे की प्रश्नावली • प्रधानाध्यपक, टीचर और स्कूल अवलोकन के लिए प्रश्नावली
1:00 PM	1:30 PM	4	• दिन के अंत में क्या करें ? • एक विद्यालय में तीनो दिन की कार्यवाही के बाद क्या करें ?
Lunch Break			
2:30 PM	4:30 PM	5	inter rater reliability exercise
4:30 PM	5:00 PM	6	FI Quiz
5:00 PM	5:30 PM	7	FI Quiz - Discussion and clarification

Table 10 : Sample Agenda for Training Workshop on Assessment Administration

8.2.3 Strategies for Effective Training

Master Trainers (MTs) should adopt the following strategies to ensure the effectiveness of the training:

1. **Session Planning:** Prepare a structured session outline, which includes introductions, recap of key points, delivery of the main training content, live demonstrations and summary of key takeaways. This ensures that all essential aspects of the training are covered systematically.

2. **Attendance and Tracking:** Maintain participant attendance daily, ensuring consistent participation. FIs should be asked to fill out the FI information sheet on both days. Those who do not attend the entire two-day training will not be allowed to go on field.

3. **Division of Responsibilities:** Assign specific roles to each MT for different segments of the session. This approach ensures that each topic is covered comprehensively and that the flow of the session remains smooth and organized with each MT focusing on their area of expertise.

4. **Time Management:** Designate a Single Point of Contact (SPOC) to manage time effectively during the training sessions, coordinating breaks for tea and lunch.

5. **Demonstrations and Practice:** Conduct live demonstrations of assessment administration to provide FIs with real-world examples of how to apply the protocols. Allow FIs to engage in hands-on practice followed by constructive feedback and guidance to ensure they are well-prepared for actual fieldwork. This interactive component reinforces learning and builds confidence in using the tools effectively.

6. **Active Engagement:** Encourage interactive discussions and Q&A sessions throughout the training. This allows FIs to clarify doubts, share experiences and deepen their understanding. Active participation ensures that FIs are not passive recipients of information but are actively involved in the learning process.

7. **Feedback and Reflection:** Incorporate regular opportunities for feedback during the training. After mock assessments or practice sessions provide constructive feedback to each FI focusing on both strengths and areas for improvement. Additionally, allow time for FIs to reflect on their performance, identify gaps and discuss strategies for improvement.

8. **Peer Learning and Collaboration:** Facilitate peer learning by encouraging FIs to work in pairs or small groups. This fosters collaboration, helps share best practices and promotes a supportive learning environment. Group activities and collaborative problem-solving can also lead to richer learning experiences and reinforce the material.

9. **Real-World Scenarios:** Introduce case studies or real-world examples that illustrate common challenges FIs may face in the field. This helps them understand how to apply theoretical knowledge in practice and equips them with problem-solving strategies for difficult situations. Some examples of such scenarios are given below.

Category	Scenario	Description
School Environment Issues	School Authority Interference	School staff push to test specific children, not as per the sampling method.
	Permission Letter Not Available / Resistance	Delay or denial of assessment due to lack of official letter or resistance from school authorities.
	Noise and Distractions	High noise levels disrupt assessment focus.
	Teacher Intervention	Teacher tries to assist or scold the child during assessment.
	Space Constraints	Limited private space for one-on-one assessment.
	PTM / Function	Assessments disrupted due to ongoing parent-teacher meeting or school event/function.
Technical Challenges	App Functionality Issues	App lags or crashes during assessment.
	Data Entry Errors	Errors while entering student details, impacting data quality.
	Inability to Sync Data	Poor network connectivity prevents timely data syncing.
	Power / Charging Issues	Phone battery drains during assessment; no charger or electricity available to recharge.
Logistical Challenges	Travel Delays	FIs delayed due to remote locations or poor roads.
	School Closed Unexpectedly	FI arrives to find school closed (holiday/strike).
	Monitoring Visit by Official	Ongoing monitoring visit may restrict or delay the assessment process.
	Pressure of Daily Targets	FIs feel rushed to meet the target number of assessments per day, risking data quality or protocol lapses.
Child-related Challenges	Child Anxiety or Fatigue	Children feel nervous or tired, affecting performance.
	Insufficient Children for Sampling	Fewer children present in school than required for sampling.
	Child Leaves During Test	Child leaves in the middle of the test, affecting the completeness of the data.
	Child Does Not Sit for All Subjects	The same child does not sit for assessments in all subjects, leading to incomplete data.
	Child Uncomfortable / Not Speaking	Child is unwilling or unable to respond, affecting assessment reliability.
	Child is differently abled	The child has speech, auditory, sight or cognitive challenges.

Table 11 : Real World Scenarios and Challenges Encountered During Data Collection

10. **Continuous Monitoring and Adjustments:** Throughout the training, monitor the progress of participants and adjust the content or delivery methods if necessary. This could involve revisiting difficult topics, providing additional examples or offering one-on-one support to those who may be struggling. Flexibility in the approach helps ensure that all FIs grasp the essential skills needed for successful data collection.

11. **Motivation and Reinforcement:** Regularly highlight the importance of the work FIs will be doing, emphasizing how their efforts will contribute to the overall goals of the assessment. Providing motivation and reinforcing the significance of their role fosters a sense of purpose and commitment to the training.

8.2.4 Final Selection of FIs

Following the 2 or 3-day training workshop, it is crucial to evaluate the participants to ensure that the data collection process is both reliable and accurate. The selection process can be done through 3 formal evaluations - quiz, observation, inter-rater reliability exercise.

1. **FI Quiz:** A quiz is conducted at the end of the second day including questions about personal details, logistical details, sampling and student collection, assessment administration protocols, data accuracy and quality check and specific assessment task scenarios. It also checks if the FI has a clear understanding of the data collection app.

   Purpose of the quiz:

   - Assess the field investigators' grasp of assessment procedures, protocols and guidelines.
   - Identify any gaps in their understanding that may need further clarification.
   - Ensure their familiarity with the use of the data collection app for data entry.
   - Reinforce the importance of standardized and ethical data collection practices.

2. **FI Observation Sheet**

   - **Ability to accurately and efficiently administer FLN.** The selected FI must demonstrate a high degree of skill in administering the FLN Assessment. This includes knowledge of administration rules and procedures, ability to accurately record students' response and the ability to use the data collection app. Assessors must be able to manage multiple tasks at once, including listening to the student, scoring the results and recording the data.
   - **Ability to establish a positive rapport with students.** FIs should be able to establish a rapport with students and help them perform the assessment to the best of their abilities. It is imperative that assessors are friendly and interact in a non-threatening manner with young children.
   - **Ability to convey instructions and questions to the child** while effectively managing any disruptions.
   - **Ability to effectively manage the app interface and student interaction.** FIs should be able to start the app correctly, maintain pace with student responses, use the timer efficiently without delays, seamlessly move to the next section, ensure all tasks are completed without skipping and facilitate the resume, save and upload functions for smooth assessment flow.
   - **Ability to manage assessment tasks effectively.** FIs should be able to conduct child sampling in schools, know when to administer stimulus and response sheets and follow the rules for auto-stop, nudging and marking the last item attempted in timed tasks to ensure accurate assessment results.

SLO Annual Assessment, HARYANA, 2024

		* Write the point number that need improvement			
		Reading instructions	Marking responses	Tablet use	Overall administration component
S. No.	Participant Name	1. Reads verbatim 2. Reads clearly (pitch, dialect and pace) 3. Reads at appropriate pace	1. Records participant responses accurately 2. Respects marking protocols 3. Marks efficiently (does not have to update marking) 4. Is quick with less response time	1. Can start the app correctly 2. Keeps pace with the student response 3. Uses the timer efficiently without delay 4. Moves to next section quickly 5. Does not skip tasks 6. Able to resume, save and upload	1. Knows the process of child sampling at school 2. Establishes good rapport 3. Knows when to give stimulus sheet and when to give response sheet 4. Knows rules for auto-stop, nudge and marking of last item attempted in timed tasks
1					
2					

Figure 21: Observation Tool to Evaluate FI Performance During Training

3. Inter-Rater Reliability Exercise

Inter-Rater Reliability (IRR) is the degree of similarity or commonality among different MTs or FIs scoring the same task. IRR ensures that there is a large degree of agreement between the assessors while scoring an item thereby increasing the scope of collecting good data.

The way of generating IRR data is to first play an audio or video of students answering an assessment that mimics a real-life assessment scenario. The FIs score the tasks (literacy and numeracy) considering it to be an on-field, real assessment. A trained MT also scores the same tasks (literacy and numeracy) and this becomes the 'benchmark score'. Then the MT will upload all the FI scores onto Excel and compare it with the trained MT's scores. If the FI does not perform sufficiently, then they may not participate in the on-field test administration and data collection. IRR checks during an assessor training workshop is essential before FIs are sent to the field. The state or the MTs may create separate sample test forms for this exercise.

8.2.5 Logistics to Address by the End of Training

To ensure that the transition from training to fieldwork runs smoothly, the following steps must be completed during the training:

1. **Distribution of Print Materials:** All Field Investigators (FIs) should receive the necessary print materials, including stimulus sheets and worksheets in the correct quantities and any required documentation such as a letter of permission from the authorities to conduct the assessment. This ensures that FIs are fully equipped.

2. **Technology Check:** Ensure that Field Investigators (FIs) are familiar with updating apps, syncing data and using GPS features. Verify that all devices are functioning correctly and that participants are capable of troubleshooting basic issues independently.

3. **Communication Channels:** Set up communication platforms such as WhatsApp groups to facilitate coordination during fieldwork. Use these platforms to share relevant updates, including school lists, day-wise field plans and any immediate field instructions or changes.

4. **Field Operations Briefing:** Provide FIs with a comprehensive operational briefing, covering daily schedules, field expenses, reimbursement policies and fixed pay. Include this information clearly in training presentations to ensure all FIs understand what to expect during fieldwork.

5. **Feedback Collection:** Gather feedback from FIs and other participants at the end of the training to identify areas for improvement. Use this feedback to make necessary adjustments for future training sessions, ensuring continuous improvement of the process.

By addressing these key logistical elements, the transition from training to fieldwork will be seamless, ensuring that FIs are prepared, supported and able to collect high-quality data.

8.2.6 FI Training Deck

The training deck should cover the following -

FI Training Deck

Agenda Overview:
Outline the schedule for all training days, including session timings and activities.

Assessment Pack:
Explain assessment framework, tool used, focus on literacy and numeracy skills, FI responsibilities.

Student Learning Outcome Survey:
Overview of survey objectives and key points, details on the number of schools per district and visits by each surveyor.

Sampling Guidelines:
Clearly explain the process of sampling students. Example : Select 7 students per class, ensuring equal representation of boys and girls.

Protocols:
- How to conduct assessment and procedure to be followed
- Required documents like permission letter

Task-Specific Instructions:
Detailed guidelines for each task in Hindi and Mathematics, including question format and expected student responses.

Question Types in Literacy and Numeracy:
Overview of timed and untimed questions in both subjects. Instructions on how to interpret and explain questions to students.

Understanding of Assessment Content:
- Use of stimulus sheets and worksheets
- Guidelines on how to assess students & handle special cases, like when a child refuses to participate

App Usage:
- How to download, log in, register on the app.
- Demonstrate app features, including data entry and using the timer function.
- Practice session with trainers for entering sample data.

Task Instructions for Students:
Explanation of color-coded text (blue, black, red) and how to read and communicate instructions effectively.

End-of-Day Procedures:
Steps to follow after completing assessments, including final data checks and debriefing.

Inter-Rater Reliability:
Introduction to the concept and its importance in maintaining assessment consistency.

Figure 22: Points Covered in FI Training Deck

8.3 Field Operations and Monitoring

Effective field operations are essential for ensuring the successful implementation of the assessment and the collection of reliable data. This sction outlines the key strategies and processes to manage the allocation of schools, monitor data collection in real-time and ensure the quality and consistency of the data being gathered by Field Investigators (FIs). The following strategies will guide the planning, supervision and quality assurance processes during the field operations phase to ensure smooth and efficient execution.

School Allotment Planning: Develop clear criteria for assigning schools to Field Investigators (FIs), taking into account factors such as geographical proximity and the comfort level of participants. Plan the allocation process to ensure efficient coverage and minimize travel challenges for the FIs. Additionally, define the number of schools to be allotted per FI ensuring a manageable workload while maintaining the quality of data collection.

Supervisors in the Field: Deploy supervisors to monitor Field Investigators (FIs) on-site during data collection. Supervisors will conduct regular visits to ensure adherence to protocols, provide guidance and address any challenges faced by Fis ensuring data accuracy and consistency.

Desk Review of Incoming Data: Establish a system for regularly reviewing incoming data remotely. Desk checks will allow the central team to identify any discrepancies or irregularities in real-time, providing the opportunity for timely corrective actions and ensuring that the data meets the required quality standards. Establishing a call centre to handle all field related challenges faced by field investigators can be considered.

8.3.1 Role of Supervisors :

Supervisors are essential to ensuring the accuracy, reliability and smooth execution of data collection during FLN assessments. One supervisor is assigned for every 5 to 6 Field Investigators (FIs). Their primary responsibilities include:

- **Overseeing Data Collection:** Supervisors monitor FIs to ensure strict adherence to assessment protocols and verify that all required data is accurately collected and complete.
- **Providing Support:** Supervisors assist FIs by managing relationships with school staff, guiding FIs to testing locations, and ensuring necessary print materials are replenished as needed.
- **Quality Assurance:** Supervisors use a monitoring checklist to conduct quality checks, ensure FIs meet their targets and confirm full compliance with protocols.
- **Coordination and Communication:** Supervisors collaborate with FIs, manage student replacements when necessary and communicate with the support team to resolve any technical issues.
- **Filling in as Needed:** If necessary, supervisors step in as assessors to maintain smooth operations and ensure there is no disruption to data collection.
- **Data Integrity:** Supervisors ensure the collected data is accurate and reliable, contributing to the overall success of the assessment.
- **Supporting Weaker FIs:** In the initial days of data collection, supervisors must also shadow the weaker FIs identified during training to provide additional support and guidance, ensuring consistent performance and maintaining high-quality data collection across the team.

8.3.2 Monitoring Checklist for Supervisors

The supervisor monitoring checklist is a comprehensive tool used for monitoring the Field Investigators' activities during the assessment process. It is designed to ensure that all steps in the assessment process are followed correctly and to provide feedback on the quality of data collection. This checklist helps maintain the standardization and quality of data collection during the SLO assessments, ensuring reliable and accurate measurement of student learning outcomes.

1. **Pre-Assessment Preparation:** This section checks if the FI has adequately prepared for the visit, including:
   - Has the information of the schools they have to visit and the schedule of the assessment.
   - Has a correct understanding of the process for sampling students within a classroom.
   - Uses the UDISE (Unified District Information System for Education) code.
   - Ensures all materials like the stimulus sheet, worksheet and other documents such as permission documents are ready and organized.
   - Performs sampling using the school register for each grade.
   - Ensures that their phone/device is charged and updated with the latest version of the data collection app.
2. **During the Assessment Visit:** This section monitors the actions and behavior of the FI during the school visit:
   - Provides proper information to the school headmaster and teachers about the visit and presents the letter of permission.
   - Is respectful and polite toward the teachers.
   - Builds rapport with the child before beginning the assessment.
   - Adheres to all sampling and assessment-related protocols.
   - Conducts the assessment with only one child at a time.
   - Uses the stimulus sheet and worksheet properly during the assessment.
   - Maintains a structured sequence when asking questions and provides clear instructions to students.
   - Encourages students to respond without hints or external help.
   - Handles disruptions or timing issues effectively, while also focusing on building rapport with the child.
3. **Post-Assessment Actions:** This final section ensures proper wrap-up of the data collection process:
   - Verification of saved data in the app.
   - Synchronizing the collected data with the central system before leaving the school. Additionally, the FI is responsible for collecting all printed materials, assessment sheets and any distributed resources.
   - Communication with the school's Headmaster (HM) about the assessment completion.

Overall Comments: A space is provided for additional feedback or observations made by the supervisor about the FI's performance during the assessment process.

8.3.3 Monitoring Incoming Data through Desk Review

The data collection app's backend data provides critical insights for monitoring the assessments. Creating a data dashboard and updating it regularly helps to track the number of students assessed daily, which enables supervisors to monitor progress and ensure FIs meet their targets. It also aids in identifying potential data falsification through data forensics, by analyzing the time taken per task and spotting discrepancies in timed questions. For example, unusually short durations on tasks or consistently unrealistic reading fluency scores can result in outliers, indicating possible misuse of timers by the Field Investigator (FI). This process further helps flag entries with implausibly short completion times and identify FIs who are repeat offenders, enabling targeted follow-up and corrective action against potential data fudging or poor-quality data collection.

By leveraging this data, supervisors can proactively detect discrepancies, provide targeted feedback and uphold the quality and integrity of the assessment process.

Component	Description	Checks to Perform	Action if Issue Found
Daily Assessment Tracking	Monitor the number of assessments conducted daily by each FI.	Track daily assessment counts per FI. Watch for unusually high numbers suggesting rushed data collection.	Investigate outliers; flag FIs exceeding realistic daily limits and verify the reasons provided by the FI. If the rationale does not appear valid, the Field Monitor should conduct a field visit or if malicious intent is suspected, drop the FI's data accordingly.
Duplicate Entries	Ensure each child's record is unique across all assessments.	Cross-verify using UDISE code (school ID), student registration/ roll number and name. Check timestamp and FI details to rule out false positives.	Remove confirmed duplicates after verification. If an FI repeatedly makes mistakes in entering student IDs, they should be contacted and provided with guidance for greater care in data entry. Additionally, the sampling sheet should be reviewed so that any incorrect information can be corrected and accurately filled in. Flag the FI if patterns persist.
Missing or Incorrect Data	Identify incomplete or inaccurate records, especially in demographic data and task responses.	Scan for blank fields, incomplete tasks (e.g., task stopped midway without valid reason), abandoned assessments or incorrect entries (e.g., age beyond expected range).	Flag entries for follow-up, data correction or completion as needed.
Time Taken Per Task and Overall Assessment	Ensure reasonable duration for each task (e.g., reading passages, number tasks).	Detect entries with unrealistically short or long durations. Compare the time taken per task as well as overall assessment duration per subject to expected timings.	Investigate anomalies; flag for potential data fabrication or rushed assessment.

Component	Description	Checks to Perform	Action if Issue Found
Consistency/ Pattern in Responses	Verify that answers across tasks follow logical patterns.	Look for identical or illogical patterns (e.g., all zeros or perfect scores or cases where children are unable to read letters yet can write sentences during a dictation task) at school and FI level.	Verify authenticity; flag suspicious response patterns.
Fluency and Accuracy Checks	Validate scores for realism; speed and accuracy should align with expected ranges.	Identify outliers, e.g., unusually high fluency scores in short durations, or uniform scores across multiple children.	Cross-verify against average scores; flag if suspicious.
Field Investigator (FI) Monitoring	Track individual FI performance across schools and days.	Identify FIs associated with high error rates, multiple flagged entries or implausible completion patterns.	Escalate issues by assigning FI shadowing by field monitors, providing targeted feedback or applying disciplinary measures. If persistent issues are identified, consider discarding the data collected by the FI.
Timestamp Verification	Ensure data collection timestamps align with school schedules and reasonable pacing.	Check for batch entries, implausibly fast assessment sequences or late-night data entries.	Investigate and verify with FI, school administration and GPS data to confirm the location of assessments.
Data Falsification (Data Forensics)	Apply forensic checks to detect data manipulation.	Analyze task durations, repeated anomalies and compare across FIs. Also, look for patterns in specific districts.	Escalate issues for corrective action and audit. If patterns of data falsification emerge across a district, consider recommending a repeat of the data collection cycle in that particular district.

Table 12 : Parameters to Check During Desk Monitoring Along With Suggestive Actions

9 DATA ANALYSIS AND REPORTING

9

Data Analysis and Reporting

This Chapter provides a structured approach to transforming raw assessment data into actionable insights for improving foundational literacy and numeracy. This process is critical for ensuring that the outcomes of assessments are accurately interpreted and effectively utilized by the decision-makers at the state or district level.

9.1 Data Cleaning

Data cleaning is a vital step to ensure the accuracy of assessment data collected during the rollout. It serves as a critical phase between data collection and analysis, ensuring the data's quality and reliability. Adhering to established assessment procedures and protocols during data collection is crucial for maintaining high standards. Data cleaning further reinforces this by addressing inconsistencies and errors, ensuring the dataset is ready for analysis.

9.1.1 Data Cleaning Checklist:

1. **Ensure that all Assessments are Linked with the Appropriate School Information for Identification**

   Remove any assessments that are not appropriately identified or labeled with the school name, district name, UDISE code, and any other marker that has been pre-decided. If the Field Investigator (FI) has mistakenly collected data from a school that was not part of the sample, the data from that school must be dropped to maintain the integrity of the sampling process.

2. **Exclusion of Data Recorded Outside the Data Collection Window**

   Remove any test data entries recorded before the official data collection period began. Additionally, discard any data entered into the data collection app after the data collection window closes, as this may reflect unauthorized entries by the Field Investigator. Ensure that only assessments conducted during the official data collection period, i.e., on the designated dates and within the allotted time slots, are included in the cleaned dataset for analysis.

3. Review Incomplete Assessment

Check incomplete assessments to determine the level of completeness and appropriateness to remain in the final data. Criteria for accepting/rejecting incomplete assessments should be decided beforehand. Incomplete assessments could be retained, for example, for the sample size or to compare and use data of the tasks completed, etc.

4. Checking for Duplicates

There may be instances of multiple children having the same name, especially in certain states. To accurately identify duplicate entries, it is important to use a combination of identifiers- the UDISE code (unique school code), the student registration number or roll number, and the student's name. Additionally, reviewing the timestamp of when the data collection form was filled, along with the Field Investigator (FI) information, can help detect any discrepancies or patterns that may indicate duplicate entries.

5. Handling Missing Student Information

If information such as name, social category, age or Child ID is missing, it can be cross-verified with the sampling sheet. If necessary, the surveyor or the school can be contacted to obtain the missing details.

6. Identifying and Removing Invalid Data

Data that does not meet the required criteria should be identified and removed. This includes:

- Instances of fudged data, such as assessments completed inaccurately or with fabricated responses.
- Cases where the criteria are not met, such as assessing a Grade 3 child instead of a Grade 2 child or conducting assessments for the wrong students.
- Assessments conducted at a school not included in the sampling list.

To identify such invalid data, a desk-check method should be employed to review and flag discrepancies. Indicators of fudging include unusually short time spent on tasks or inconsistencies in responses. Flagged data can be discussed with the FI for clarification and invalid entries must be dropped to maintain data quality.

7. Discrepancies Between Forms

If discrepancies are found between the literacy and numeracy forms filled by the Field Investigator, they should be manually checked and corrected. For example, if the FI incorrectly enters the student registration number in one of the forms, this can be identified during the manual data review process. Once the error is spotted, the correct information should be updated in the corresponding backend database to ensure consistency across all records.

8. Ensure that All Timed Subtask Scores Fall Within an Acceptable and Realistic Range of Scores

During data collection, errors introduced by FIs or software glitches might lead to the presence of extreme outliers in the scores. It is important to investigate any exceptionally high scores and verify that they are realistic for the pupil being assessed (by comparing their performance in other subtasks) and are not the result of some error. Any extreme observations based on flagging criteria should be removed so that the clean data is not skewed. Please note that only erroneous data is to be removed for a particular student. The rest of the data from the other tasks should be retained in order to maintain the sample size.

9.1.2 Outlier Detection and Removal

Outlier detection involves identifying data points that significantly deviate from the expected range or patterns. These outliers can distort the results and affect the accuracy of the analysis.

Scenario: In the ORF dataset, the observed fluency rates (measured as items read correctly per minute) show significant variability. Some unusually high values may be attributed to potential issues such as glitches in the data collection application, incorrect use of the timer or other unidentified factors.

Min	0
Max	3600
Mean	90.6781
Median	29
SD	292.76

Table 13. Summary Statistics from ORF Data Showing High Variability

The Interquartile Range (IQR) method for outliers is a statistical technique used to identify data points that significantly deviate from the rest of a dataset by calculating a range based on the middle 50% of the data (between the first and third quartiles) and considering any values falling outside a certain multiple (usually 1.5 times) of that range as outliers. Essentially, it defines outliers as data points that lie significantly farther away from the majority of the data than the middle 50% of the distribution.

Step 1: Use the IQR method and boxplots for outlier detection by setting the upper and lower bounds.

Outlier Detection Using the IQR Method	
Q1	11.5
Q2	29
Q3	59.01639
Q4	3600
Interquartile Range (Q3-Q1)	47.51639
Lower Bound=Q1-IQR*1.5	-59.7746
Upper Bound= Q3+IQR*1.5	130.291

Table 14. Quartile Statistics and Outlier Boundaries in ORF Data Using the IQR Method

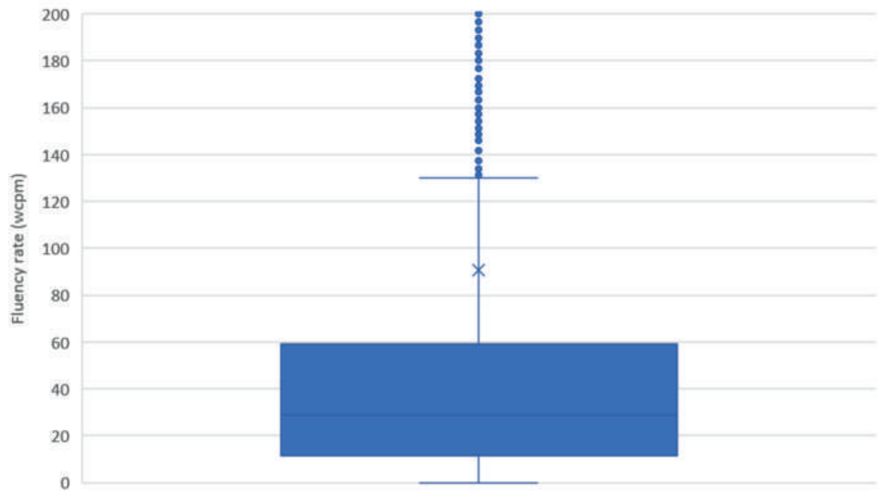


Figure 23: Boxplot Illustrating Outliers in Oral Reading Fluency Rates (wcpm)

Step 2 : The data with fluency rates above 130. New descriptive statistics are -

Min	0
Max	130
Mean	31.28037
Median	25.40541
SD	28.13687

Table 15. Summary Statistics from ORF Data Post Outlier Treatment

But now, we still see outliers shown in the below scatter plot. One case marked in square is where a child who could read only 10 words but had a fluency rate of 120 wcpm as he took only 5 seconds to read those 10 words, giving a false picture.

This could happen as we had applied a 4-word discontinuation rule where if the child reads 4 consecutive words incorrectly, we discontinue the task or due to incorrect use of the timer.

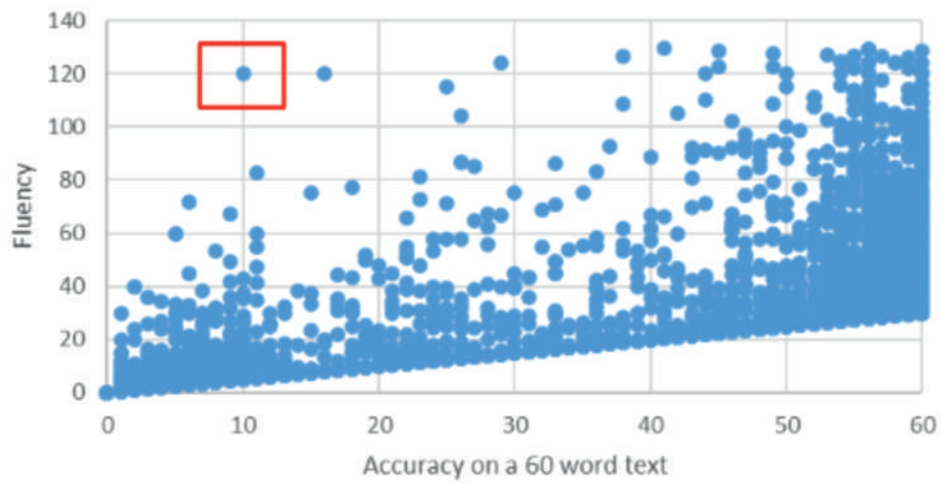


Figure 24: Scatterplot of Fluency and Accuracy on the ORF Task, Showing Remaining Outliers

Step 3: Thus, to treat such outliers as well, calculate the ratio between the fluency rate and accuracy and the records with a ratio greater than 3. The scatterplot shown below reflects the relationship between accuracy and fluency in the ORF task for the new dataset that remained after dropping such outliers.

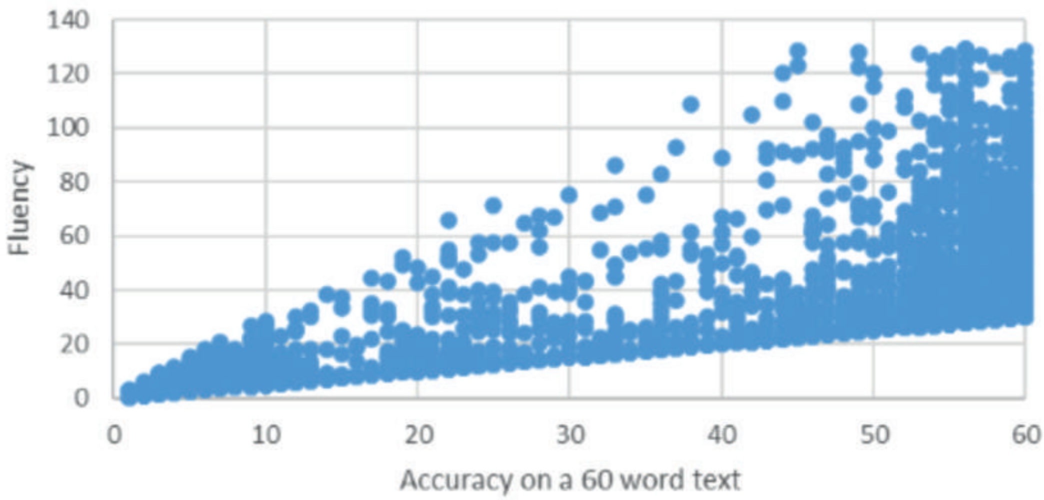


Figure 25: Scatterplot of Fluency and Accuracy on the ORF Task, After Complete Outlier Treatment

The new and final descriptive statistics for the ORF fluency rates is given below -

Min	1
Max	129
Mean	33
Median	27
SD	27.55

Table 16. Final Summary Statistics from ORF Data Post the Complete Outlier Treatment

9.2 Data Analysis

Once the data cleaning procedures have been completed, it is ready for analysis. In the FLN data analysis phase, researchers use descriptive and/or inferential statistics to describe the data, examine patterns and draw conclusions. Each of these methods has a different purpose and value.

9.2.1 Descriptive Statistics (Non-Inferential)

Descriptive (or non-inferential) statistics are used to describe and summarize data keeping in mind the objective of seeing **what patterns might emerge.** This helps in presenting the data in a meaningful way that can be easily interpreted. The most common measures reported in descriptive analyses are frequencies, measures of central tendency (e.g., means and medians) and measures of spread (e.g. standard deviations and summary ranges).

Descriptive statistics are used only to **describe sample data.** As discussed in the previous sections, samples are selected to be representative of larger populations. Descriptive statistics are to be reported only for studies that are designed to not conclude samples; or as unweighted frequencies, unweighted means, etc., for complex assessment data. Descriptive statistics cannot be used to conclude the data collected. Lastly, Descriptive statistics are used to fully describe the sample according to the various levels of disaggregation that have been captured and need to be reported. For FLN assessments, the data could be disaggregated by language, gender, rural vs. urban schools, tribal vs. non-tribal districts and any other parameter of disaggregation decided by the state that needs to be reported. While reporting FLN data, descriptive statistics include the means of fundamental demographic characteristics of the sample, as well as means across subtasks for every level of disaggregation.

- **Weighted Average:** A weighted average is a calculation that assigns varying degrees of importance to the numbers in a particular data set. It can be more accurate than a simple average in which all numbers in a data set are assigned an identical weight. A weighted average ensures that the aggregate results fairly represent the performance of all sampled schools and students, even when the number of students assessed varies across schools.

  In FLN assessments, both schools and students within schools are sampled. However, the number of students assessed in each school can differ. A simple average treats all schools equally, regardless of how many students were assessed in each school. This could skew the results, as schools with fewer students would have an equal influence on the overall average as schools with more students.

  A weighted average adjusts for this imbalance by assigning greater importance (weight) to schools or groups with more students assessed, ensuring that the final average reflects the true performance of the sampled population

$$\textit{Weighted Average} = \frac{\textit{Sum of Weighted Terms}}{\textit{Total Number of Terms}}$$

- **Mean:** During the assessment, a child is scored on each subtask that is administered. The mean score is the average score of a group or cohort of students on a given subtask. This means it gives a good overall snapshot of how a group of students is performing at a particular reading skill. Mean scores provide useful insight into the progression of reading skills in a group of students over time. However, mean estimates are sensitive to extreme values or can be affected by outliers like very high or very low scores, such as the number of students scoring zero.

$$m = \frac{\textit{Sum of the Terms}}{\textit{Number of Terms}}$$

- **Zero Scores:** Sometimes children are not able to answer a single question correctly on a given task. For example, the student may not be able to identify a single letter sound or read a single word. This results in a task score of zero. Zero scores are referred to as the proportion of students who cannot perform even one item correctly for a given task. A high proportion of zero scores will cause the distribution of scores to skew right (positively skewed data) and lower the mean score.

The tasks in the assessment are divided into 2 types - timed and untimed. Untimed tasks focus on assessing accuracy alone while timed tasks are used to assess fluency, combining both speed and accuracy to measure a student's performance.

- **Accuracy:** Accuracy measures the percentage of items answered correctly in an assessment by a student. It indicates how well the student performed in terms of correctness, without considering the time taken.

  Example: Calculating Accuracy

  - Number of correct items: 15
  - Total number of items: 20

$$\textit{Accuracy} = \left(\frac{\textit{Number of correct Items}}{\textit{Total Number of Items}}\right) X\ 100$$

$$\textit{Accuracy} = \left(\frac{15}{20}\right) X\ 100 = 0.75\ X\ 100 = 75\%$$

So, the accuracy on the task is 75%.

- **Fluency :** Fluency is the ability to read text quickly, accurately and with proper expression.

EGRA's most direct measurement of fluency, the oral reading fluency with comprehension task, is a core component of the instrument. Children are given a short written passage on a familiar topic and asked to read it out loud "quickly but carefully." Fluency comprises speed, accuracy and expression (prosody). The oral reading fluency task is timed and measures speed and accuracy in terms of the number of correct words read per minute. This task does not typically measure expression.

Fluency is measured for timed tasks and calculated in items per minute.

Example: Calculating ORF Fluency

- Number of words read correctly: 120
- Total time taken to read (in seconds): 60

$$\text{Fluency} = \left(\frac{\text{Number of correct Item}}{\text{Time Taken (in seconds)}} \right) X \ 60$$

$$\text{ORF Fluency} = \left(\frac{120}{60} \right) X \ 60 = 2 \ X \ 60 = 120 \ \text{wcpm}$$

So, the student's oral reading fluency is 120 words correct per minute (wcpm).

9.2.2 Inferential Statistics

Unlike descriptive statistics, evaluators can use inferential statistics to conclude entire populations based on a sample, infer hypotheses about population parameters and compare different populations (e.g., treatment and control groups). These statistics are crucial for both impact evaluations and FLN assessments that aim to make statements about education in a state or district based on a sample of students or schools. The selection of inferential statistics for a particular student depends on the research design as outlined below:

	Baseline					Endline							
Treatment	Mean fluency (wpm)	Standard error	Number of sampled students	*t*-stat	*p*-value	Mean fluency (wpm)	Standard error	Number of sampled students	*t*-stat	*p*-value	Difference–in–difference	*p*-value (DID)	Effect size
Control	4.5	0.6	656	–	–	9.5	1.6	475	–	–	–	–	–
Intervention	5.2	1.2	349	0.510	0.611	11.7	1.1	480	1.189	0.236	1.5	0.490	0.12

Calculations:

Difference-in-Difference: (Mean endline treatment – mean baseline treatment) – (mean endline control – mean baseline control)

Effect Size (Cohen's *d*): Difference-in-difference / pooled standard deviation

Table 17. Example in Difference-In-Difference Analysis

9.2.3 Regression Analysis

Regression analysis is the most common method for analyzing variables and predicting their relationships and it can be used in the context of FLN assessments as well. Ordinary least squares (OLS) regression is suitable for FLN data with normally distributed residual values, especially when a continuous variable like oral reading fluency score is utilized. However, in many developing countries, test scores are clustered around zero, leading to an uneven score distribution. In such cases, evaluators may consider binomial regression analysis, such as probit or logistic regression, to examine binomial outcomes like whether a student meets local reading ability benchmarks or scores zero on a specific reading subtask.

9.2.4 Analysis of Background Questionnaires

A student's background can influence their learning. Factors like parental education, family income and home environment shape their educational journey. This section outlines the analysis of background questionnaire data to understand these influences and uncover meaningful insights.

Background questionnaires often ask about the following:

- Parents' occupation and education: How educated the parents are and what they do for a living
- Family income and home setup: Household resources and infrastructure
- Home environment: Language spoken at home, learning support (e.g., helping with homework) and activities like reading practice, taking books home or watching educational videos

These questions paint a picture of the child's context and help us analyze how these factors connect to learning outcomes, such as their NIPUN status or proficiency in key skills like Oral Reading Fluency (ORF).

The goal is to understand how a child's background affects their learning. This analysis helps:

- Explore relationships between background factors and learning indicators (e.g., NIPUN status)
- Identify patterns in learning outcomes based on socioeconomic and home factors

Step 1: Clean the Data

Before diving into the analysis, the data must be cleaned:

- **Remove Duplicates:** Ensure each child ID appears only once and all children have valid IDs.
- **Label Variables:** Use clear, meaningful names for easier interpretation.
- **Standardize Values:** Clean up categorical variables to avoid inconsistencies (e.g., "high school" vs. "High School").

Step 2: Merge with Learning Outcomes Data

Combine the cleaned background data with student learning outcomes. The final dataset should include:

- The child's background information.
- Their NIPUN status (overall, literacy and numeracy).
- Scores in competencies like ORF fluency.

This merged dataset provides a complete view of each child's context and performance.

Step 3: Analyze the Data

1. Descriptive Statistics:

- Create tables and charts to summarize patterns.
- For example:
    - Compare NIPUN status with levels of parental education.
    - See how many children practice reading at home and how it relates to their performance.

2. Inferential Statistics:

- Correlations: Explore connections, like whether ORF fluency improves for children who practice reading at home.
- Significance Tests: Check if these relationships are statistically meaningful (e.g., parental education and NIPUN status).

3. Regression Analysis:

- Predict outcomes using background data.
- For instance:
    - Estimate the likelihood of a child being NIPUN-compliant based on family income or home reading habits.
    - Measure the influence of different background variables on learning outcomes.

Analyzing background data helps reveal the factors that shape student learning.

By connecting background characteristics with NIPUN status and key competencies, we can pinpoint disparities and design targeted interventions to support students who need it most. It is a step toward making education equitable and effective for every child.

9.2.5 Benchmarking

A benchmark is a standard or point of reference against which things may be compared or assessed. Creating benchmarks is essential for setting clear expectations and norms for student performance. Benchmarks provide specificity to the curriculum, ensuring clearer learning goals and measurable progress. By translating long-term objectives into manageable performance targets, they help gauge progress at specific points in time.

For educators, benchmarks are instrumental in guiding teaching efforts, enabling teachers, principals and supervisors to identify areas requiring targeted interventions. Additionally, benchmarks facilitate public communication about improvements in education through mechanisms like school report cards and national monitoring reports, promoting transparency and accountability.

Each state sets its benchmarks for both numeracy and literacy assessments to evaluate student performance.

NIPUN benchmarks represent ultimate competencies that students are expected to achieve, such as reading 40 words per minute or meeting other defined performance standards by the end of grade. A child meeting these benchmarks in literacy, numeracy or both is considered NIPUN. These benchmarks are analyzed at state, district and block levels, rather than at the school level since not all schools are included in the sample. This prevents drawing generalized conclusions about individual schools.

District rankings based on NIPUN performance can help identify areas with the highest and lowest percentages of NIPUN students. Such insights are crucial for resource allocation, allowing states to prioritize low-performing areas and direct efforts where they are most needed.

9.3 Data Presentation

Data presentation helps in effectively communicating the results of assessments to diverse stakeholders. Clear and engaging data presentation is essential to ensure that findings are understood and utilized for decision-making. This section emphasizes the use of visual tools like graphs and maps to summarize complex data, making it accessible to policymakers, educators, and program implementers. Different types of charts include -

1. **Proficiency Bands:** Proficiency bands are a visual tool used to categorize and analyze student performance across various tasks. They break down performance into distinct levels (scores or rates), making it easier to understand where students excel or struggle. The image below shows the proficiency levels for both accuracy and fluency tasks.

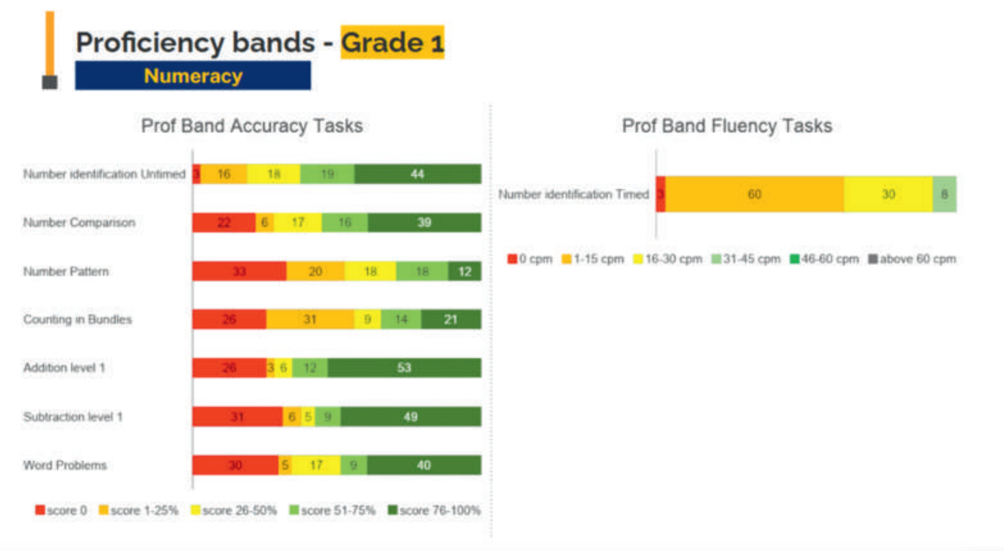


Figure 26: Representation of Proficiency Bands

- **Accuracy Tasks:** These proficiency bands help to identify the percentage of students in different performance levels like, 0 scores, 1-25%, 26-50%, etc. for tasks like number identification, addition, etc. This allows educators to pinpoint areas where a significant number of students are underperforming, such as word problems or number patterns.

- **Fluency Tasks:** These proficiency bands categorize students' performance based on their fluency that is the correct responses per minute. Here, the data highlights how many students are in lower bands, 0 wcpm versus higher bands above 45/60 wcpm, offering insight into the spread of fluency across the cohort.

2. **Bar Charts:** These are mainly used to compare different groups.

   - **Comparison of zero scores:** Bar graphs are a useful tool for comparing data, such as how students performed on the same tasks in previous years versus the current year. They can also be used to compare metrics like the proportion of zero scores. Once the data is visualized, trends become easier to identify, such as a decrease in the number of zero scores over time.

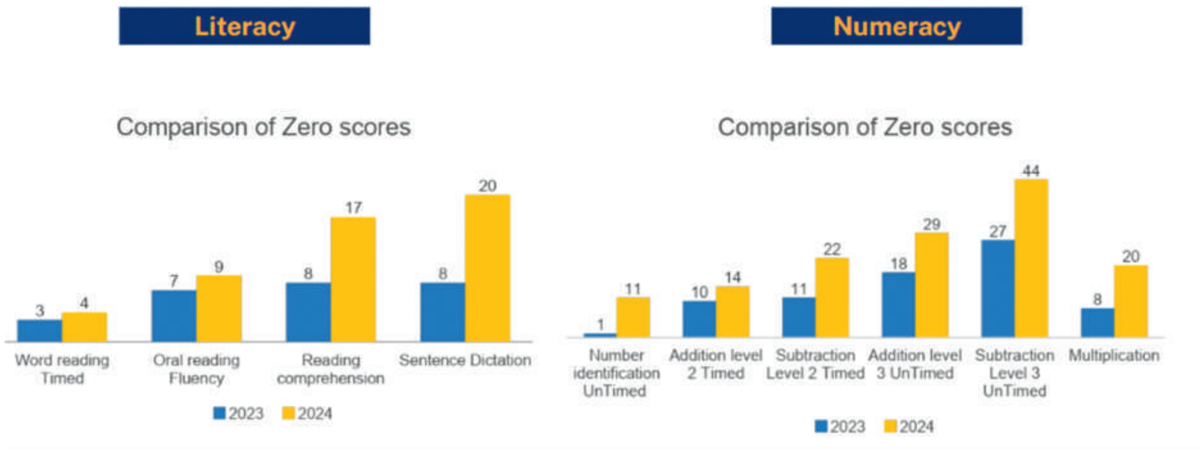


Figure 27: Comparison of Zero Scores Through Bar Charts

- **Comparison of correct answers on comparable items:** Bar charts are also helpful for comparing students' performance on comparable items over the years. For example, 73% students correctly read "बतख" (duck) in 2023, whereas only 65% students read it correctly in 2024. This helps track changes in student performance over time, providing insights into learning progress and areas of stagnation or decline.

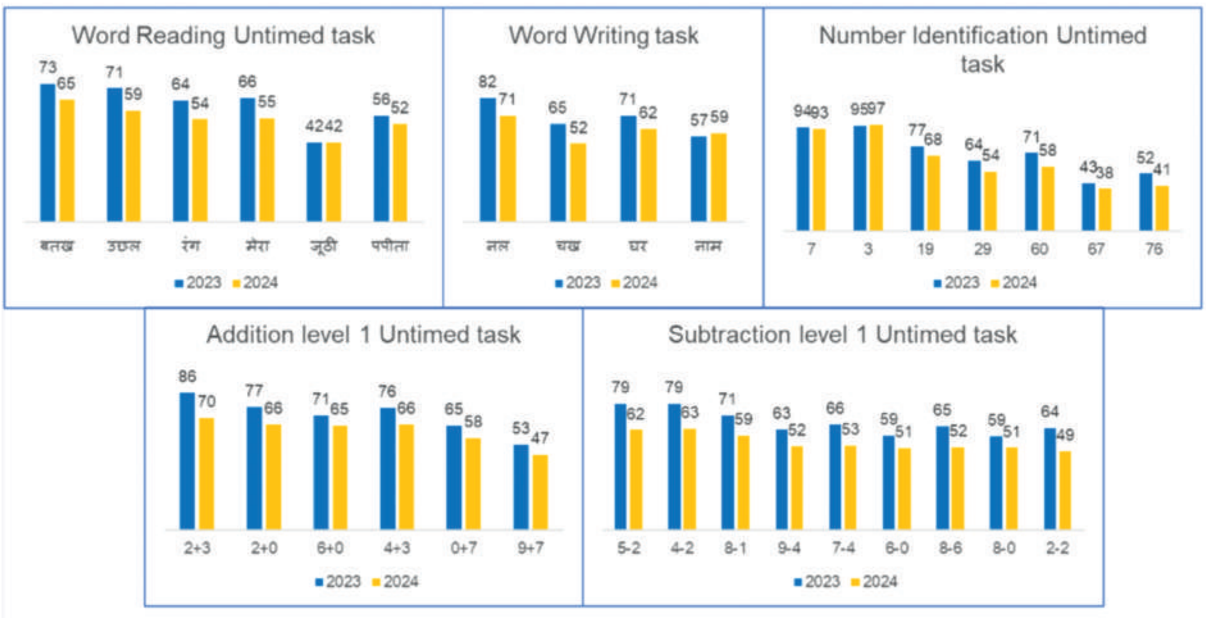


Figure 28: Comparison of Student Performance Through Bar Charts

- Bar graphs can be used to compare average scores over the years for various tasks and to illustrate the % of students who have passed the benchmark score.

3. **Scatter Plot:** A scatter plot identifies a possible relationship between changes observed in two different sets of variables. It provides a visual and statistical means to test the strength of a relationship between two variables.

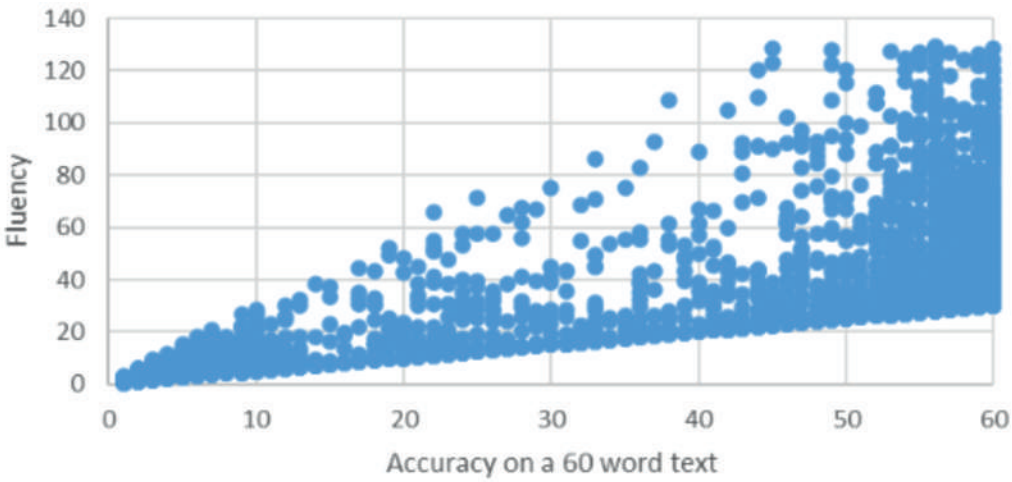


Figure 29: Scatterplot of Fluency and Accuracy on the Timed Task

4. **Heat Maps:** Heat maps are utilized to show the performance levels of different regions, with color gradients indicating the varying degrees of success. These maps help quickly identify which districts are performing well and which require attention, providing a clear visual representation of data at a glance.

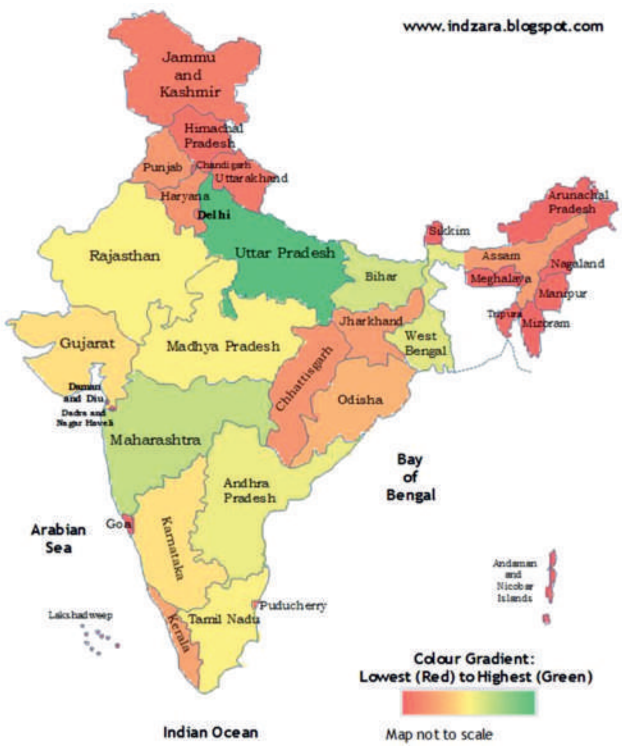


Figure 30: Heat Maps Showing Performance Levels of Different Regions

9.4 Usage of Data

Data from the assessment can be used in the following ways :

- Analyze gaps in reading competencies to raise awareness and drive improvements in policies, curricula and instructional practices at the national or regional level.
- Identify key skills and areas of instruction that need to be improved to target interventions such as teacher training and resource development.
- Support strategic planning within the education sector by informing resource allocation, budgeting and decision-making processes.
- Track changes and trends in literacy and numeracy outcomes over time to monitor progress and identify persistent challenges.
- Assess the effectiveness of programs aimed at improving early-grade reading or numeracy skills through impact or performance evaluations.
- Establish clear reading indicators and benchmarks to measure and monitor literacy or numeracy achievements effectively.

10 DISSEMINATION AND USAGE OF DATA

10

Dissemination and Usage of Data

Assessment data can only drive change when it is shared purposefully and reaches the right audiences. Without a clear dissemination strategy, valuable findings may remain underutilized. A well-planned approach helps ensure that data informs policy, improves teaching and builds public awareness around foundational learning.

10.1 Planning for Effective Dissemination

An effective dissemination strategy outlines what will be shared, with whom, how and when. This includes developing communication materials, setting timelines, assigning responsibilities and identifying necessary resources. A structured plan helps ensure that key messages are communicated clearly and used meaningfully.

The following key steps can guide the development of such a plan.

1. Define the Objective of Dissemination Strategy

- The objective should be aligned with the objectives set for conducting the large FLN assessments.

2. Identify Key Target Audiences

- The target audience should be aligned with the objectives defined.
- A specific objective can have multiple audiences.
- A suitable product aligning with the target audience should be identified.

3. Decide on a Timeline for the Dissemination Strategy

- The timeline can include preparing and engaging the target audience with some follow ups if required.

4. Identify and Tailor Key Messages and Reporting Product

- The key message should be aligned with the audience and objectives of the assessment.
- It should focus more on actionable information, interventions and follow-up.
- The reporting product should capture the key messages in simple and understandable language.

5. Identify and Customize Communication Materials and Communication Channels

- The material developed should be in an accessible format with infographics.
- Multiple communication channels including online and offline modes can be used.

6. Provide an Ongoing Plan or a Follow-Up Plan

- Continue to engage with stakeholders by sharing practices, and webinars, etc for follow-up and establish a feedback mechanism with the stakeholders.

Figure 31 : Key Steps in Developing a Dissemination Strategy

10.1.1 Objectives of Dissemination Strategy

The objective of a dissemination strategy is to ensure that assessment findings are communicated clearly, purposefully, and in a timely manner to support informed decision-making and meaningful action. The strategy should align with the intended use of the data and be tailored to the needs of different stakeholder groups.

Common objectives include:

- **Inform Policy and Planning:** Provide policymakers with evidence to guide reforms, develop new initiatives or make informed decisions about resource allocation.
- **Support Curriculum and Teacher Development:** Use assessment insights to revise curricula and design targeted teacher training that addresses specific learning gaps.
- **Improve Classroom Practice:** Help educators identify learning gaps, adapt instructional strategies and monitor student progress to enhance teaching and learning.
- **Promote Accountability and Transparency:** Share findings widely to foster shared responsibility and encourage continuous improvement among schools, districts and education departments.
- **Engage and Empower Stakeholders:** Keep stakeholders, including educators, parents and community members informed and involved through accessible data and opportunities for dialogue.
- **Raise Public Awareness:** Highlight the importance of foundational learning and build broader support through public campaigns, such as FLN Day.
- **Strengthen Community Participation:** Encourage parental involvement and local partnerships to create a more supportive learning environment for children.

10.1.2 Identifying Key Stakeholders

Effective dissemination requires a clear understanding of who needs the information and why. The target audience must be aligned with the objectives of the dissemination strategy. A single objective may correspond to multiple stakeholder groups, each with different informational needs and levels of influence.

- **Align Audience with Objective:** For example, if the goal is to inform classroom practice, the primary audience might include teachers, school leaders and teacher trainers. If the objective is to drive policy change, then government officials and education planners would be key stakeholder.
- **Map Multiple Audiences to Each Objective:** Recognize that each dissemination goal may require engaging several audiences and messaging should be tailored accordingly.
- **Select Suitable Products for Each Audience:** Different stakeholders require different formats—briefs for policymakers, data dashboards for administrators, or visual summaries for community groups. Choosing the right product enhances clarity and usability.

A well-mapped stakeholder plan ensures that findings are delivered in the right format, to the right people, at the right time. The following figure illustrates the key stakeholder groups at the state, district, block, school and community levels.

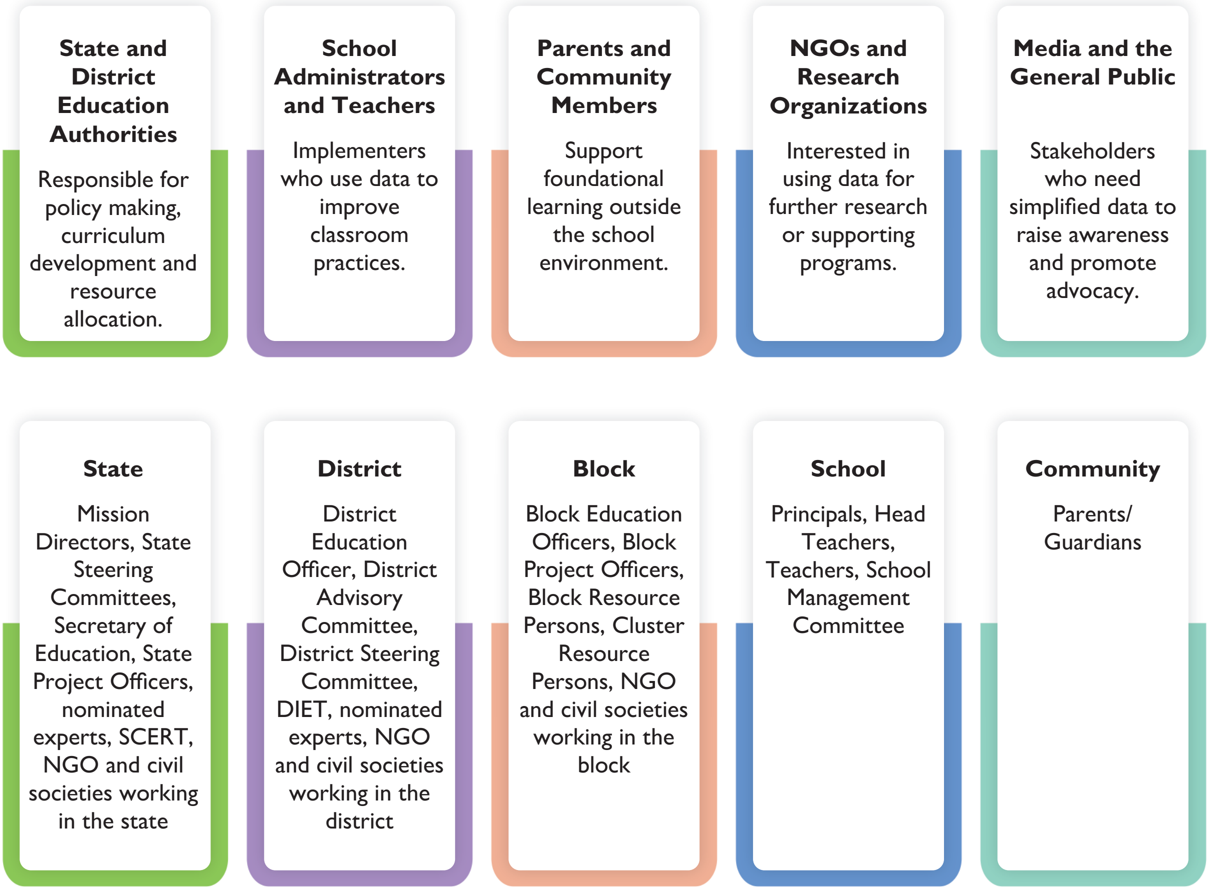


Figure 32: Key Stakeholders for Dissemination at Various Levels

10.1.3 Decide on a Dissemination Timeline

Establishing a clear timeline is essential to ensure that findings are shared in a timely and coordinated manner. The timeline should align with the decision-making cycles of different stakeholder groups and allow sufficient time for stakeholders to process the findings, respond and take necessary actions.

- **Plan for Timely Release:** Schedule dissemination to coincide with relevant planning windows—for example, academic year cycles, budget planning or teacher training programs.
- **Sequence Communication:** Consider a phased approach, starting with internal validation, followed by targeted sharing with key decision-makers and then broader public dissemination.
- **Allow Time for Action:** Build in adequate time between dissemination and follow-up activities so that findings can be meaningfully acted upon, whether through policy decisions, training initiatives or school-level interventions.

A well-timed dissemination strategy increases the chances that the data will not only be seen, but also used.

10.1.4 Identify and Tailor Key Messages and Reporting Products

Clear, concise messaging is critical for effective dissemination. Key findings should be translated into messages that resonate with each stakeholder group, emphasizing relevance and actionable insights.

- **Tailor Messages to Audience Needs:** Different stakeholders have varying levels of technical expertise and interests. For example, policymakers may require high-level summaries with implications for policy, while teachers might benefit from practical tips linked to classroom practice.
- **Highlight Actionable Insights:** Focus on what stakeholders can do with the information—whether it's adjusting strategies, allocating resources, or increasing community engagement.
- **Choose Appropriate Reporting Formats:** Select communication products that best suit the audience and purpose. These could include detailed reports, policy briefs, infographics, presentations, dashboards or video summaries.
- **Ensure Clarity and Accessibility:** Use simple language, visuals and examples to make findings understandable and engaging, especially for non-technical audiences.
- **Choose Appropriate Reporting Formats:** Select communication products that best suit the audience and purpose. These materials may include:
    - **Executive Summaries:** Brief documents summarizing key findings for busy policymakers.
    - **Main Reports:** Comprehensive analyses for education experts and decision-makers.
    - **Technical Reports:** Detailed accounts for researchers and technical experts.
    - **Policy Briefs:** Concise documents for decision-makers, highlighting critical findings and recommendations.
    - **Media Reports and Press Releases:** Simplified versions for the general public to raise awareness and advocate for change.
    - **Thematic Reports:** Focused on specific issues, such as gender disparities or regional differences in foundational learning outcomes.

Executive Summary

Content: Key findings, summary of initial data

Audience: All stakeholders, media, public

Detail: Low

Channels: Websites, print, email, blogs

Main Report

Content: Full assessment overview

Audience: All stakeholders, media, public

Detail: Medium

Channels: Websites, print, email, workshops

Technical Report

Content: Detailed assessment data

Audience: Key stakeholders, researchers

Detail: High

Channels: Website, print

Assessment Framework Report

Content: Definitions, context, reporting

Audience: Stakeholders, researchers, public

Detail: Medium to High

Channels: Website, print

Thematic Report

Content: In-depth topic analysis

Audience: Specific stakeholders, researchers

Detail: Medium to High

Channels: Print, website, seminars

Policy Briefings

Content: Summary and implications

Audience: Policymakers

Detail: Low to Medium

Channels: Print, PPT/videos, infographics

Media Reports / Press Releases

Content: Public information and advocacy

Audience: General public

Detail: Low

Channels: Newspapers, blogs, videos, infographics

Assessment Database

Content: Clean data

Audience: Researchers, officials

Detail: High

Format: .csv / .xlsx

Figure 33: Various Types of Reporting Formats

- **Choose the Right Dissemination Approach for Each Stakeholder Group:** Dissemination efforts are most effective when communication formats and methods align with stakeholder preferences and contexts:
    - **State-Level Stakeholders:** Use formal presentations, policy briefs and workshops to communicate implications for policy and program direction.
    - **District-Level Stakeholders:** Organize district meetings, host community forums and leverage local media to ensure broad understanding and buy-in.
    - **School-Level Stakeholders:** Conduct school-based dissemination through teacher workshops, staff briefings and parent-teacher conferences to connect findings directly to classroom practice.

For example: In one of the states, district-level dissemination involved partnering with local radio stations to broadcast key findings from the FLN assessment, thereby engaging parents in remote areas.

Customizing messages and formats ensures the assessment data is meaningful and usable across different levels of the education system.

10.1.5 Identify and Customize Communication Materials and Channels

To maximize the reach and impact of dissemination efforts, it is essential to select and tailor both communication materials and dissemination channels according to the preferences and needs of different stakeholder groups.

- **Selecting Appropriate Dissemination Channels:** Different audiences access information through varied channels, so it's important to choose the most effective ones for each group. Dissemination channels can be broadly categorized into:
    - **Digital Channels:** Including websites, webinars, emails, social media platforms and digital reports—ideal for timely and wide-reaching communication.
    - **Offline Channels:** Such as workshops, printed reports, community meetings and face-to-face presentations—effective for interactive engagement and detailed discussions.
    - **Mass Media Engagement:** Utilizing press releases, radio, television and infographics to raise awareness and reach a broader public audience.
- **Mixing Channels for Maximum Impact:** Employing a combination of digital, offline and mass media channels ensures that the findings reach stakeholders in formats and platforms they trust and regularly use, thereby enhancing engagement and the likelihood of meaningful action.
- **Ensuring Ongoing Stakeholder Engagement:** Engagement should not be limited to the moment of dissemination but should continue throughout the process. This creates opportunities for stakeholders to provide feedback, participate in discussions and contribute to decision-making. Effective engagement methods include:
    - **Workshops and Conferences:** Facilitate discussion of findings and the sharing of best practices.
    - **Formal Briefings and Presentations:** Inform senior officials and policymakers about critical results and actionable recommendations.

By combining tailored communication products with ongoing engagement, dissemination efforts can promote better understanding, ownership and use of assessment findings across all levels.

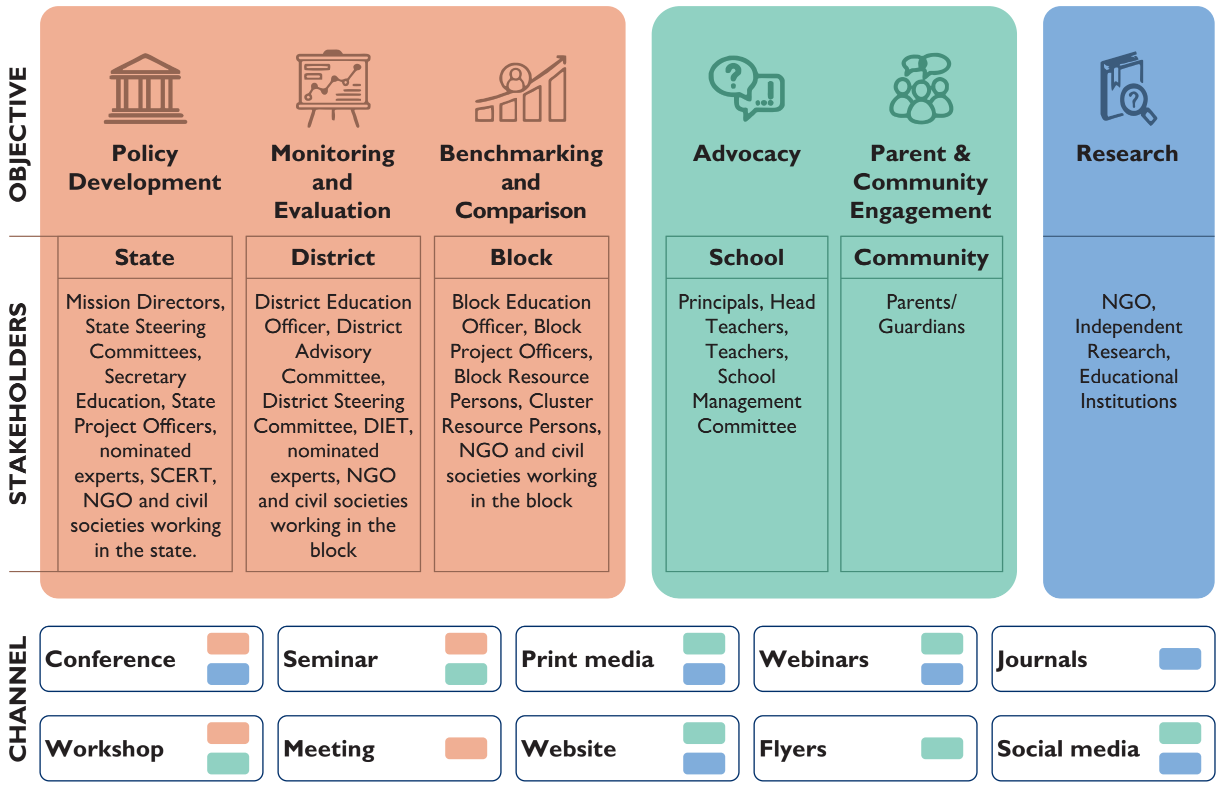


Figure 34: Mapping Dissemination Objectives, Stakeholders, and Channels

The figure below illustrates the key objectives of dissemination, the relevant stakeholder groups and the appropriate communication channels for each. For example, at the state, district and block levels, dissemination may aim to support policy development, enable monitoring and evaluation, or facilitate benchmarking and comparison of student learning outcomes. Suitable channels at these levels include conferences, seminars and focused meetings with key officials.

10.1.6 Provide an Ongoing Follow-up Plan: Setting Targets Using Assessment Results

Sustained impact from assessment data requires more than just initial dissemination; it demands ongoing follow-up and clear target setting to guide improvement efforts.

Target Setting

Setting realistic and measurable targets based on assessment results is essential to drive progress in foundational learning. Targets create clear goals and benchmarks, helping educators and policymakers track improvements over time and stay focused on priority areas.

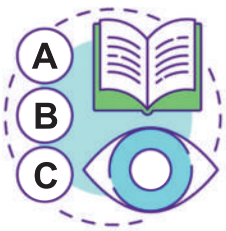
Steps to setting targets :

- **Analyze Assessment Results:** Review key trends, strengths and gaps revealed by the data.
- **Establish Baseline Performance Levels:** Use current data to set baseline levels for critical skills such as reading and mathematics.
- **Identify Key Areas for Improvement:** Pinpoint specific skills, grades or student groups needing targeted support.

- **Set Realistic and Achievable Targets:** Define goals that are Specific, Measurable, Achievable, Relevant and Time-bound (SMART).
- **Communicate Targets and Expectations:** Clearly share targets with all stakeholders, clarifying roles and responsibilities in meeting them.
- **Develop a Monitoring and Evaluation Plan:** Implement a system to regularly track progress, assess effectiveness and adjust strategies as necessary

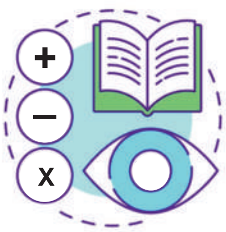
A Sample of Target Setting Using State Foundational Learning Assessment Results

Reading skills



Baseline performance level	Target setting	Interventions and strategies
Average fluency rate - 35 correctly read words per minute	35 wcpm → 2 years → 45 wcpm	• Provide targeted professional development for teachers on evidence based reading instruction techniques • Implement a structured reading program that focuses on building reading fluency • Introduce supplementary reading materials and resources for students to practice reading at home

Numeracy skills



Baseline performance level	Target setting	Interventions and strategies
45% students could correctly solve more than 75% of the number operation questions	45% → 2 years → 70% *Ambitious but achievable targets*	• Provide targeted professional development for teachers on effective strategies for teaching basic arithmetic operations • Implement a math program that emphasizes hands-on learning and problem-solving • Encourage parents to engage in math related activities with their children at home

Figure 35: A Sample of Target Setting Using State Foundational Learning Assessment Results

10.1.7 Monitoring and Evaluation of Dissemination Activities

In addition to tracking learning progress, it is important to monitor the dissemination strategy itself. Regular monitoring and evaluation help assess the effectiveness of dissemination activities. Key indicators can include:

- **Reach:** Number of stakeholders engaged through each dissemination channel.
- **Engagement:** Participation rates in events such as workshops, webinars and meetings.
- **Impact:** Observable changes in policy, teaching practices or awareness that result from dissemination efforts.

This ongoing process ensures that assessment findings translate into actionable steps, driving continuous improvement in learning outcomes.

Case Study: Target Setting and Academic Planning

A sample-based baseline assessment of Grade 2 students across all districts in a state covered 48,000 children from 1,600 schools. It provided a clear picture of foundational literacy and numeracy levels, highlighting areas for urgent action.

To act on these findings, the state initiated a structured process of **academic planning and target setting** aimed at improving FLN outcomes over five years. Six rounds of three-day workshops were held with ~240 participants, including officials from SCERT, DIETs, assessment cells and teachers. A technical team supported workshop design, materials and facilitation.

Workshop Agenda:

- Day 1: Dissemination of district-wise analysis of assessment data.
- Day 2: Literacy target setting and identification of related strategies.
- Day 3: Numeracy target setting and planning.

Example: State-Level Target for Oral Reading Fluency (ORF)

At baseline, only 12% of students were reading at an "Above Benchmark" level (35–54 correct words per minute). The state set progressive annual targets:

Year	Target (% students Above Benchmark)
2022–2023	25%
2023–2024	40–45%
2024–2025	55–60%

Each district developed its own targets based on reflections on current resources and limitations.

Planning and Reflection Using FLN Targets Worksheet

Districts used a structured worksheet for intervention planning aligned to their FLN targets. It included:

- Current status
- Proposed strategies
- Responsibility
- Timeline

Planning covered five key areas:

1. **Critical Enablers:** Teacher availability and attendance, TLM delivery, infrastructure
2. **TLMs:** Language kits, math resources, digital content
3. **TPD and Mentoring:** Training plans, NISHTHA modules, mentoring systems
4. **Assessments:** Student Based Assessment (SBA), dipstick assessments, state-level surveys
5. **Monitoring Systems:** KPIs, data systems, governance mechanisms

Districts adopted different approaches—some phased targets gradually, while others set ambitious early goals. Interventions were planned to align with indicators, though some required support linking actions to outcomes.

BIBLIOGRAPHY

- Ball, M.-C., Curran, E., Tanoh, F., Akpé, H., Nematova, S., & Jasińska, K. K. (2022). Learning to read in environments with high risk of illiteracy: The role of bilingualism and bilingual education in supporting reading. Journal of Educational Psychology, 114(5), 1156–1177. https://doi.org/10.1037/edu0000723
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. Journal of Economic Growth, 17(4), 267–321. http://www.jstor.org/stable/23324889
- Ministry of Education, Government of India. (2020, July 29). National Education Policy, 2020 [PDF]. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
- Ministry of Education, Government of India. (2021). NIPUN BHARAT: National Initiative for Proficiency in Reading with Understanding and Numeracy. [PDF] {Link: https://www.education.gov.in/sites/upload_files/mhrd/files/nipun_bharat_eng1.pdf}
- Scanlon, D. M., Gelzheiser, L. M., Vellutino, F. R., Schatschneider, C., & Sweeney, J. M. (2008). Reducing the incidence of early reading difficulties: Professional Development for classroom teachers versus direct interventions for children. Learning and Individual Differences, 18(3), 346–359. https://doi.org/10.1016/j.lindif.2008.05.002
- Torgesen, J. K. (2002). The prevention of reading difficulties. Journal of School Psychology, 40(1), 7–26. https://doi.org/10.1016/S0022-4405(01)00092-9

ANNEXURE I

LISTENING COMPREHENSION

The child has to listen to a story read out by the assessor and then orally answer 4 comprehension questions based on the same story.

TASK DEVELOPMENT

Story

- Create third person narrative of a story with 1/2 characters with only 1 proper noun.
- Make original or new story — not similar to common stories or legends with which children maybe familiar.
- Have a logical flow with a beginning, a problem/confrontation and a resolution to it.
- Do not include a story title or illustrations.
- Avoid violence or other themes that might be troubling to children.
- Use positive gender representations.
- The entire story should be in the same tense.
- The length of the story should be grade appropriate.
- The words used should be grade appropriate in terms of frequency, length, number of syllables.
- Have grade appropriate grammar and sentence structure (avoid conjunctions and dependent clauses).
- Avoid including dialogue with quotation marks.

Questions

- Three should be explicit (answer found directly in text).
- One should be inferential (answer not found directly in the story).
- The questions should refer to the story, but do not need to be distributed evenly or aligned to sequential ideas.
- Include the correct answers next to the questions for easy reference for assessors.

ORAL VOCABULARY

The child has to name 6 objects shown in the form of a picture one by one.

TASK DEVELOPMENT

- The objects chosen for this task should be from everyday vocabulary of students and appropriate to grade level.
- The names of objects and verbs should be garde appropriate. For e.g., eating is appropriate for Grade 1 while driving is for grade 2/3.
- Corresponding pictures chosen should be clear, relevant and free from any kind of bias - cultural, gender, social.
- Pictures should have a thick black outline for ease of children with visual difficulties.
- Choose black and white images.
- Images should not be cartoonish.
- Could include common objects, body parts, professions, vegetables, fruits, animals, etc. and common actions.

PHONEMIC AWARENESS (INITIAL SOUND IDENTIFICATION)

The child has to identify initial sounds in 10 words spoken to them one by one.

TASK DEVELOPMENT

- Select simple 1 - 2 syllable words, frequently occurring in grade level textbooks.
- These words should not be used in any other task in the same tool.
- Categorize the words according to initial sound type and point of articulation. For e.g., in English
    - Vowel: a, e, i, o, u
    - Fricative (voiced or unvoiced): v, f, s, z, th, sh, j, x
    - Plosive: b, p, d, t, k, g
    - Nasal: m, n
    - Semi-vowel, approximant, or other: h, y, w, r, l
- Keep at least one word representing each type.
- Don't keep words with the same initial sounds e.g., bag, ball.
- Don't keep words displaying similar consonants e.g., broke, choke.
- Don't keep words with blends (i.e., "st" or "pl") as initial sound for grade 1.
- Have one (and only one) word with an initial vowel in the first five words.
- Have one (and only one) word with an initial vowel in the last five words.
- For grades 2 and 3 you can ask some questions on last sound recognition as well.

PICTURE COMPREHENSION

The child has to answer questions based on a composite picture.

TASK DEVELOPMENT

- Use bright, colorful, engaging and simple illustrations.
- Ensure the scene aligns with familiar environments (e.g., parks, schools, markets).
- Avoid clutter; the picture should have distinct, easily identifiable elements.
- The picture should have distinct characters, doing distinct actions (easily identifiable).
- Represent diversity in characters and activities.
- Create 3 direct and 1 infer level item.
- Use clear, simple language appropriate for Grade level.
- Relate questions to obvious, observable elements in the picture.
- E.g., of direct fact retrieval item - What is the girl doing?
- E.g., of infer level item - Why do you think the children are happy?
- Begin with the easiest question and increase complexity gradually.
- Ensure each question has a clear, logical answer based on the image, avoiding subjectivity in answers.
- Avoid confusing questions e.g., if there are two girls in the image doing different activities, don't ask " what is the girl doing?" instead be specific "What is the girl wearing cap doing?"

LETTER NAMING

The child has to recognize 50 letters.

TASK DEVELOPMENT

- Determine how many unique graphemes (smallest functional unit of a writing system) there are in the alphabet (e.g., 26 for English).
- Choose practice letters that demonstrate different types of letters.
- Place the letters in a grid with 10 columns and 10 rows.
- Write each letter once, in random order, in a mix of upper and lower case letters.
- Within the remaining spaces: place each letter, in random order, in the opposite case from how it appears the first time. [for English]
- For other languages, put letters with matras as per the blueprint.
- If more spaces in the grid remain empty, enter, in random order, the letters proportionally to how their frequency in textbooks.
- Ensure that each letter in the alphabet appears at least once before a letter is repeated.
- Ensure that the first letter in the grid is not the first letter of the alphabet or a practice letter.
- Ensure that the order of letters does not result in a familiar acronym or in an obvious word spelled within the grid.
- Ensure that identical letters do not appear next to each other. Eg., b and d

FAMILIAR WORD READING

The child has to read 50 familiar words.

TASK DEVELOPMENT

- Develop a word frequency list including all parts of speech from grade textbooks.
- Prepare a grid (empty table) with 5 columns and 3 rows for untimed task and 5 columns and 10 rows for timed task.
- For untimed task, choose 15 familiar words from word frequency list, for timed task, choose another 50.
- Do not repeat words from the untimed task in the timed task.
- Depending on the frequency list, choose a suitable proportion for nouns, verbs and adjectives.
- Word length and spelling patterns are representative of those found in early grade readers (i.e., consonant – vowel – consonant, for English words; one - and two -syllable words).
- Words are composed of a variety of letters, with none repeated disproportionally.
- None of the words has alternative or ambiguous pronunciation (for example, “read” or “record” in an English language test).
- None of the items is a word in any other language to which the children may be familiar.
- None of the words is specific to a subregion or dialect.
- Items do not include one - letter words, as these will already be included in the letter grid.
- Enter items in random order (NOT in order of difficulty, length, alphabetic order, etc.) in a 10 row * 5 column well spaced grid.
- Choose two additional words to use as example words for assessors to practice with students. The words should be similar in level of difficulty to the words in the grid.

NON WORD READING

The child has to read 50 pseudo words or invented words.

TASK DEVELOPMENT

- Choose 1–3 frequent orthographic patterns appearing in grade 1 textbooks (i.e., CVC or CVCV).
- Create non-words that follow these frequent orthographic patterns.
- Use a range of letters to create the nonwords; do not use any consonant disproportionally to others.
- All words should be nonwords (i.e., they do not have any meaning in the language assessed or another language with which the children may be familiar), in written and spoken form. For example, "kat" would not be an appropriate nonword in English because it is a homophone of "cat."
- Nonwords should follow the rules of the language, using letters in legitimate positions (e.g., in English, cannot have "wuj" because "j" is not used as a final letter in English).
- Word lengths and spelling patterns should be representative of those found in beginning readers (i.e., consonant – vowel – consonant).
- Enter items in random order (not in order of difficulty, length or in the same order as the familiar words) in well spaced 10 row * 5 column grid.
- Develop two additional nonwords to include as example words and update the directions.

SENTENCE READING

The child has to read 5 individual sentences.

TASK DEVELOPMENT

- Create 5 individual non-connected sentences.
- Use the words from the frequency list.
- The sentences should not exceed the average sentence length in the textbooks.
- Ensure sentences align with Grade 1 vocabulary and comprehension levels.
- Arrange sentences progressively, starting with the simplest (shortest) and gradually increasing in complexity.
- Ensure the transition feels natural and not abrupt.
- Start with simple subject-verb-object constructions (e.g., "The cat jumps.").
- Progress to sentences with modifiers or conjunctions (e.g., "The big dog runs fast.").
- Avoid tricky phrasing.
- Use relatable and engaging context.
- Vary subjects and actions to make the task engaging.
- Do not repeat an article disproportionately.

ORAL READING FREQUENCY

The child has to read a story aloud , quickly but carefully.

TASK DEVELOPMENT

- Create a third person narrative around a story of 1 or 2 characters with one proper noun only.
- The story should be original or new—not similar to common stories or legends with which children may be familiar.
- Avoid character names frequently used in textbooks.
- The story must have a logical flow with a beginning, a problem/confrontation and a resolution to it.
- Do not include a story title or illustrations.
- Avoid violence or other themes that might be troubling to children. Use positive gender representations.
- The entire story should be in the same tense.
- The length of the story should be 40-60 words only.
- The words used should be grade appropriate in terms of frequency, word length, number of syllables.
- The story should have grade appropriate grammar and sentence structure (avoid conjunctions and dependent clauses).
- Avoid including dialogue with quotation marks.

READING COMPREHENSION

The child has to answer 5 questions based on the story read in the oral reading fluency task.

TASK DEVELOPMENT

- Three should be explicit (answer found directly in text).
- Two should be inferential (answer not found directly in the story).
- The first question should correspond to the first sentence (or approximately 7–10 words).
- Develop remaining questions to evenly correspond to text (for example, one question every 10–15 words).
- For inferential questions regarding the story overall, note that the child would need to read at least 80%-100% of the story.
- Questions should be complete statements and not "fill in the blank": "Where did the boy go?" instead of "The boy went to the ___?".
- Avoid questions with multiple correct responses. For example, ask: "What did the girl buy at the market?" (Answer: "rice" or "bread") rather than "What are two things the girl bought at the market?"
- Avoid having questions that a child could correctly answer without having read the story. "What color was the apple?" (answer: red)
- Answers should not require too much subjective interpretation by assessors to determine whether they are correct.
- Include the correct answers next to the questions for easy reference for assessors.

LETTER DICTATION

The child has to write 5 letters dictated to them.

TASK DEVELOPMENT

- Include both capital and small letters (e.g., A, b, D, e, etc.) or
- Include letters both with and without matras (e.g., क, की, म, मा).
- Ensure a balanced mix of easily distinguishable letters (e.g., no visually confusing pairs like 'b' and 'd' in the same set).
- Begin with simple, frequently used letters (e.g., क, म).
- Gradually introduce akshars with matras or more complex formations (e.g., की, तो).
- Arrange the 5 letters in increasing complexity to challenge students step by step.

WORD DICTATION

The child has to write 5 words dictated to them.

TASK DEVELOPMENT

- Use high-frequency words from the textbook.
- These words should not be tested in any reading task.
- Choose words that are phonetically regular and familiar to students.
- Include a mix of 2-letter, 3-letter, and gradually longer words (e.g., "it," "cat," "sand").
- Words are composed of a variety of letters, with none repeated disproportionately.
- Start with shorter and simpler words.
- Gradually increase complexity by introducing longer words or those with blends and digraphs (e.g., "jump," "train").
- Avoid highly complex or rarely used words.
- Arrange the 5 words from simple to complex.
- Ensure the words are within the expected phonetic and comprehension level of students.
- Use meaningful and contextually appropriate words.

SENTENCE DICTATION

The child has to write 5 sentences dictated to them.

TASK DEVELOPMENT

- Use sentences made from high-frequency words and structures in the textbook.
- These sentences should not be tested in the sentence reading task.
- Also avoid words that have appeared in any of the reading task.
- Words are composed of a variety of letters, with none repeated disproportionally.
- Select sentences with familiar contexts.
- Keep sentences short and simple, especially for early learners (e.g., 4-6 words).
- Gradually increase length while maintaining clarity.
- Begin with basic subject-verb-object structures (e.g., "The dog runs.").
- Progress to sentences with adjectives, conjunctions, or possessives (e.g., "The red ball is big.").
- Begin with very short and phonetically regular sentences.
- End with the longest sentence of the set, using a variety of sentence elements. Examples: "The little girl is reading a big book."
- Ensure avoiding commonly mispronounced words.

ANNEXURE II

PRE NUMERACY SKILLS - PRE-MEASUREMENT

The child is able to apply the knowledge of pre numeracy skills (length, weight, capacity) in different situations in an untimed task.

TASK DEVELOPMENT

- Identify the seven situations/concepts for pre numeracy skills.
- Design a situation based on the different skills identified. The measurement has to be happened between two objects or three objects (wherever it is mentioned in the curriculum).
- The situations should be familiar with to the students .
- The skills should be known to children and are in increasing order of complexity. There should be one situation for each of the skill identified.
- Ensure the font used is such that it does not make two numbers appear alike.
- Ensure the pictures used should be clear and feasible for a grayscale printout.
- Label the objects (A/B/C) or (i/ii/iii) and not (1/2/3).

NUMBER IDENTIFICATION (UNTIMED) UPTO 500

The child is able to recognize 10 numbers given in a numerals form in an untimed task.

TASK DEVELOPMENT

- Create a table of 2 rows and 5 columns.
- The numerals should be of increased difficulty:
    - The first six items of the subtest include the numerals 0, 9 and other four two-digit numbers upto 99.
    - The next 4 items consist of three digit numbers from 100 to 500.
- Font used should be such that it does not make two numbers appear alike.
- The numerals should be familiar to the student and all the numerals should be present atleast once.
- Students are instructed to say the numbers out loud.

NUMBER IDENTIFICATION TIMED UPTO 99

The child has to recognize and verbally identify 20 numbers with increasing complexity in a timed task. (60 seconds)

TASK DEVELOPMENT

- Make a table of 4 rows and 5 columns.
- The numerals should be arranged in increased difficulty.
- The first four items of the subtest include the numerals 0, 9 and two other single-digit number.
- The next 16 items consist of two-digit numbers from 10 to 99. (9 upto 50 and 7 upto 99).
- Font used should be such that it does not make two numbers appear alike.
- The numerals should be familiar to the student and all the numerals should be appearing twice at least.
- Numerals should be aligned with the LO framework. Give the situations of mirror image, place value etc.
- Students are instructed to say the numbers out loud.

NUMBER COMPARISON (UNTIMED) UPTO 500

The child has to make judgements about differences by comparing 10 pairs represented as numbers.

TASK DEVELOPMENT

- Make two columns, each column will have 2 set. Each item should contain two numbers, one should be greater than the other.
- The first item is a set of one-digit numbers, the next five items are sets of two-digit numbers up to 99 and the last four items are three-digit numbers up to 500.
- Clear instructions should be provided as to whether the student needs to identify the larger number
- The situations should be familiar with to the students
- Items should not contain equal numbers.
- The value of items should be jumbled so that larger and smaller items appear on both left or right columns
- Items should appear in the order of 1 one-digit numbers, then 5 two-digit items up to 99 and then 4 two-digit items up to 500.

COUNTING IN BUNDLES

The child is able to apply the knowledge of counting in bundles to given situations in an untimed task (5 questions)

TASK DEVELOPMENT

- Identify the five situations for counting in bundles (1 questions from end of grade 1 and 4 from grade 2 (if it is being taught till Oct)
- Design a situation based on the different situations identified in increasing order of difficulty. First question less than 9, 2nd and 3rd less than 50 and 4 and 5 upto 99.
- The situation should be broken down into different levels, starting with basic counting and moving towards counting up to 99.
- The situations should be familiar with to the students. Counting using matchsticks pictures (bundles of matchsticks/or something which is used in the textbooks)
- Ensure that the numbers asked in this question are again not asked in the Number identification question
- Ensure the font used is such that it does not make two numbers appear alike.
- Ensure the pictures (bundles and sticks) used should be clear and feasible for a grayscale printout.

NUMBER PATTERN IDENTIFICATION

The child should be able to determine the missing number in a pattern of four numbers.

TASK DEVELOPMENT

- Make four horizontal columns, three containing numbers and one which is empty.
- Based on the grade level , the numbers in the pattern can be increased by ones, fives, tens etc.
- Make multiple such items of varying difficulty levels keeping grade appropriateness and LOs in mind.
- The number pattern should be consistent, i.e. if it has a gap of fives, all the numbers should appear in gaps of fives.
- The blank spaces should be provided in such a manner that it is not always at the end of item.
- Multiple items should be made of varying difficulty keeping grade appropriateness in mind.
- The instructions should be clear for the students to solve the number patterns.
- The students should be familiar with the numbers and patterns.

ADDITIONAL LEVEL 1 AND 2

The child is able to solve addition problems with sum below 20 (level 1). Apply the procedural knowledge to solve two digit numbers involving bridging (level 2).

TASK DEVELOPMENT

- Ensure that the addends (numbers to be added) and sum (total of the addends) are grade appropriate.
- Write two numbers horizontally with a plus sign followed by an equals sign and a box or a dash to fill the answer.
- In the instruction mention, that the numbers are to be added and the solution is to be written in the box or blank provided
- The task should cover all the possible situations for addition like addition of two digit with two digit, addition of two digit with single digit etc.
- The sum of the addition should also cover all the possible situations like sum ending with 5, 0 etc.
- The resultant sum should be according to the grade level and the curriculum when the assessment is planned. Eg., if the assessment is planned in September, the academic calendar has an addition of numbers up to 50.
- Tally the addends and the sums with the LO to keep the item grade appropriate.

SUBTRACTION LEVEL 1 AND 2

The child is able to solve subtraction problems with sum below 20 (Level 1). Apply the procedural knowledge to solve two digit numbers involving bridging (level 2).

TASK DEVELOPMENT

- Ensure that the numbers to be subtracted are grade appropriate.
- Write the larger number is on the left and the smaller number is on the right.
- Write two numbers horizontally with a minus sign followed by an equal's sign and a box or a dash to fill the answer.
- In the instruction mention, that the numbers are to be subtracted and the solution is to be written in the box or blank provided
- The task should cover all the possible situations for subtraction like subtraction of two digit with two-digit, subtraction of two digit with single digit etc.
- Check the numbers to ensure they are aligned to the grade level Los.

WORD PROBLEMS

The child has to apply addition and subtraction procedural skills to word problems.

TASK DEVELOPMENT

- Identify the type of word problem which will be the most grade appropriate and level specific based on the LO. Ensure that the problem is testing only a numeracy LO and not a literacy LO .
- Write the word problem statement using two or three sentences in clear language. Use the context for framing the word problem.
- The numbers below ten should be expressed in words and not in numerals - (two and not 2)
- The question should be open ended, i.e. there should not be options like an MCQ .
- The statements should be unbiased.
- Ensure that the word problem does not use any unfamiliar words.
- Ensure that the word problem requires a clear number operation (addition or subtraction)
- Word problems can be of the following types :

Problem Type	Example
Change: Result Unknown	Two children are on the bus, three more children get on. How many children are on the bus altogether?
Combine: Result Unknown	There are six children on the bus, two are boys. The rest are girls. How many girls are there on the bus?
Compare: Change Unknown	There are two children on John's bus and seven children on Mary's bus. How many children must join John's bus so that it has the same number of children as Mary's bus?
Change: Start Unknown	Five children get on the bus. Now there are 12 children on the bus. How many children were on the bus to begin with?
Sharing	Four children share twelve candies equally between themselves. How many candies does each child get?
Multiplicative	There are five seats on the bus. There are two children on each seat. How many children are on the bus altogether?

SHAPE RECOGNITION

The child is able to recognize specific shapes amongst the collection of shapes.

TASK DEVELOPMENT

- On a sheet create four to five 2D and 3D shapes in standard form, ensuring that the shape being tested appears more than once in varying length, height, width or radius - depending on the shape.
- Create 7-8 non-standard shape forms, open shapes, shapes with curved boundaries.
- Create 7-8 shapes in the surroundings matching the shapes.
- The instruction should clearly mention which shape with a standard form among all these shapes should the child identify.
- Ensure that the shapes are familiar to the students and are level appropriate.
- Ensure that there is a good mix of standard and non-standard shapes.
- Ensure that the standard and non-standard form shapes are jumbled such that all the standard form shapes are not concentrated on one side.

ANNEXURE III

Example: Planning number of FI and days required for assessment

Suppose 60 schools in a district need to be assessed.

- **Option 1:** Hiring 60 FIs ensures that each FI visits one school. This minimizes fieldwork time but significantly increases costs due to higher training and personnel expenses.
- **Option 2:** Hiring 30 FIs, with each visiting two schools, reduces costs but requires more days in the field, increasing logistical complexity while partially addressing budget concerns.

Additional considerations:

If an FI is assigned to visit multiple schools across different grades, the schedule may look like this for each school:

- **Day 1:** Assess Grade 1 students in all subjects
- **Day 2:** Assess Grade 2 students in all subjects
- **Day 3:** Assess Grade 3 students in all subjects

The decision regarding the number of FIs and the number of days required per FI must carefully balance time, cost and logistics:

- **Increasing the number of FIs** reduces the duration of fieldwork but raises training and operational costs.
- **Optimizing the number of visits per FI** helps distribute the workload more efficiently, potentially reducing costs and managing logistical challenges.

Design and Artwork: ReseTricks Research and Data Consultancy Services (contact at: resetricks@gmail.com)

CENTRAL SQUARE FOUNDATION

To know more about CSF

Visit our webiste

www.centralsquarefoundation.org

Follow us on social media

CSF_India

central-square-foundation

CentralSquareFoundation

central_square_foundation

CentralSquareIndia