

This is the last section of a report of a large scale school based assessment of student achievement conducted in May 2014 by Bihar government (SCERT), ASER Centre/Pratham and UNICEF.



Lessons and learnings from the May assessment exercise in Bihar:

Some concluding thoughts

The assessment exercise carried out in May 2014 jointly by Bihar Government, ASER Centre/Pratham and UNICEF has been a big source of learnings. There are lessons in it for the way forward not only for how to organize and how to conduct future assessments but also for what can be done to improve teaching practices and learning outcomes.

Large scale assessments as opportunities for capacity building

To begin with, collaboration between the three participating organizations meant that the strengths of each of the partners could be brought into the project. Cross fertilization of ideas and partnership implies that the learnings from the process could be absorbed more easily by all parties not just from the findings but also from the process of carrying out the exercise. During the design and decision making phase, 20-25 people were actively involved; these were state level officials of the government and senior people from ASER Centre/Pratham and UNICEF. Once the basic structure and content was in place, a team of 200 + people from the government and from ASER/Pratham led the three week effort from start to finish. Then the focus of activities moved to the districts. Here 2500 surveyors (DIET students and Cluster Coordinators) spent almost a week in over 1000 schools across the state. The teams interacted with approximately 65000 children. Despite the heat of the summer and the pre-monsoon storms, one of the unique things about this initiative was the energetic participation of a large number of people (adults and children) from within the education system and outside.

Large scale assessments of student achievement are fast becoming a common feature of the academic calendar in all states. Often the main activity remains limited to being a data collection effort. One of the major lessons from the experience in May 2014 in Bihar is that such an assessment exercise can be an excellent platform for capacity building and hands-on learning. The big challenge is how to productively use the time that becomes available during these large scale assessments for maximizing the potential of human interactions (adults with children, children with children and adults with adults) to improve our understanding of how children can learn better.

Rapport building with children: The design of the May assessment was such that each team of two surveyors/evaluators spent several days continuously in a school. For two outsiders, going to a school daily for several days is a good way to get to know the school and the children. In addition, we also wanted to make sure that all activities were taking place in a non-threatening environment. As warm-up activities for rapport building, we trained all surveyors to play simple group games with children – some games were based on language skills and some on maths. A booklet called “Aao Khelein” was given to all participants which contained many such games and ideas.

Solving problems together: After the written tests were done, many children wanted to know how they had performed. A simple activity that was both effective and fun was to solve each question from the test paper together with children. For example, children in Std 6 discussed the question, both what had to be done and how to do it and then solved it or came up with the answers. This was sometimes done individually and sometimes in groups. In a very direct way, the benefit of such activities is that the children can learn immediately from the assessment tasks. This interaction also gave the test administrators an opportunity to interact with children around teaching-learning.

Discussions with teachers: The presence of the assessment team (as well as visits by the master trainers) in a school for several days in a row naturally led to many conversations about assessment, instruction, curriculum and learning. Although there is meant to be ongoing school based activities on assessment such as CCE and child wise report cards, there is usually not much of an opportunity to have on-going discussions about why to measure, what to measure, how to measure and or about ways to convert the learnings from the measurement into actual action at the school level. In future, assessment teams could be oriented with some guidelines for content and structure about how to initiate and sustain such discussions with teachers. Also, due to the fact that many children were absent, there were extra blank test papers available after the assessment was done. Often, teachers wanted to keep these papers as examples that they could use both to create similar exercises and also to use with children. This can be a suggestion that is given to future assessment programs as a way to use extra material.

Capacity building built into the assessment activity: Throughout the May exercise, there was a conscious attempt to integrate capacity building about assessments into ongoing activities wherever possible. For example, during the state level training of master trainers, there were sessions that introduced some of the basic concepts of assessment.

The team of surveyors at the district level were either DIET students or Cluster Coordinators. Both such groups are either already closely linked to the life of schools or are likely to become so in the very near future. During the district level training also, efforts were made to introduce some of the nuts and bolts of assessment practices – such as how the assessment framework was constructed and why methods need to be implemented systematically and consistently. In the future, more such sessions could be included as part of district-level training.

Further, each day after school time was over, all the surveyors along with their master trainers gathered together in a convenient place (cluster or block resource centres, or in the DIET or district headquarters) to discuss the day's proceedings and to grade the papers. All the written tests were graded by the surveyors for the schools assigned to them under the supervision of the master trainers and based on the guidelines for grading that had been provided to them. Grading papers helped the evaluators and the master trainers to see common patterns of mistakes and weaknesses and also to link experiences (of the class/school they were visiting) with evidence (what children had been doing). In the stray incidents of copying, this close look at the papers the same day as the assessment helped to

sort out what was to be done. In future, such exercises can also maximize learnings from using these daily sessions better than was done in this round.

Assessment courses and dissemination workshops: Now that this state level student assessment exercise is done, the findings and the lessons learned are a good starting point for a basic level course on measurement that can be conducted at the state level. Such discussions are currently on between ASER Centre/Pratham, the state government and UNICEF. Also as series of workshops are being planned for district level teams as well as for Cluster Coordinators to disseminate the findings from the assessment exercise. Participation from each district and DIET would mean that capability for carrying out useful such exercises would increase across the board along with the ability to translate evidence into action.

Overall, the participatory nature of the May assessment opened the doors for building capacity. We believe that feedback loop at every level is critical for the evolution, development and improvement of any activity. This work in Bihar shows that assessments can be designed with feedback built in – feedback to children via solving questions together and feedback to teachers via discussions. Our experience also suggests that data and findings make much more sense to all stake holders, especially those at the field level when there has been participation in the entire process. When evaluators/test administrators have understood the framework, implemented the assessment, interacted with the assessed children and graded papers, they are in a much better place to appreciate findings. Ownership is also much greater when there is direct engagement in the implementation and when key people are not just the passive recipients of a report. That is why, who participates and how in such exercises is an important consideration for future use of data.

Lessons for future assessments:

The almost month long immersion in assessment activity meant that a large number of people were constantly and actively thinking about the entire process. Key thoughts and lessons are outlined here.

Importance of reading as a fundamental and foundational skill: The Bihar May 2014 assessment is perhaps the first, recent large scale state level assessment that included the assessment of basic reading as a core part of the exercise along with the more common practice of using pen-paper/written tests. As is well known, if a child cannot read fluently and comprehend, the chances that she or he will succeed in the education system is low. Hence reading is one of the most fundamental skills that needs to be built in the primary grades. The reading assessment in the May exercise has led to several major learnings:

- *High incidence of children who cannot read fluently:* There are substantial number of children in Std 4 and even in Std 6 who have difficulty in reading Std 2 level text fluently. We need to think about how to help these children “catch up” with others.
- *Reading assessment can only be carried out individually one-on-one:* Most assessments in India like NCERT’s National Achievement Surveys or those done by other agencies – are done with groups of children. However to assess children’s reading ability, there is only one way to do it: work with children one by one. This is more time consuming and needs more training to make sure that the assessment is

being done consistently and systematically. However to really understand children's reading levels there is no other way that is possible.

- *Availability of data makes a problem visible and its solution possible:* The availability of data on reading (for example – how many children in which grade cannot read) makes it possible for the government to make plans and programs for how children's reading skills can be strengthened. Without such data, remedial or learning support activities cannot be planned. In the past, despite the availability of data from the annual ASER reports, reading has not been directly tested in government surveys. Without measurement by the government, the crisis in reading has remained invisible. Now with government data becoming available for primary and upper primary grades, it becomes possible to design and carry out activities that help to solve this problem.
- *Inability to read affects performance in maths:* The data generated in this exercise shows how the inability to read strongly influences the performance of children in maths as well. Other subjects were not assessed in May but it is likely that the ability to read is strongly correlated with how children cope with content in other subjects.
- *If a child cannot read, she or he cannot do pen-paper written tests:* This is common sense and data from May strongly confirm this statement. Almost all large scale student achievement surveys are entirely based on written tests. As more and more states do large scale assessments of the pen-paper kind, this fact should be kept in mind. Administering written tests to children who cannot read excludes information about their abilities from the evidence and makes it unlikely that their problems will be the focus of attention. There are equity implications of such exclusion. In future assessments, a test of basic reading could possibly be used as a "screener" before the child is asked to take a written test.

Are we testing curriculum or are we testing children: In designing assessment frameworks for any exercise, it is important to be clear about their purpose. If the objective is to see whether how children's performance compares with curricular expectations of a particular grade then the domains and items should be based on grade level curriculum standards. However if the objective is to understand what children can do and what they cannot do, then the framework needs to have tasks that range in difficulty from very easy to difficult. If most items are above the current level of most children, then the assessment will not be able to provide useful inputs for what needs to be done to help all children improve.

Available data for India suggests that a large fraction of children at every grade are several years below their grade level. Hence it is imperative that large scale assessments developed for use in Indian elementary schools have a diverse range of tasks so that useful evidence can become available for use in planning appropriate teacher training programs, and teaching learning activities and materials for children. Such actions will be needed to take children on the journey - from the level at which they are, to the level at which they need to be.

Children are not familiar with multiple choice formats: In the May assessment, care was taken before every written assessment to explain how multiple choice questions had to be handled. There were examples in the test paper and there were repeated demonstrations with children. Yet the analysis of the mistakes for the questions in the written papers shows

that a significant section of children did not understand how to handle the multiple choice format. Multiple choice formats are the most convenient form for grading however depending on the context and age of children, they may not be the best for children.

How quickly do children get tired? Theory of testing suggests that multiple items are needed to get at the true estimate of a child's competency level. However, tests also should be designed keeping in mind how long an average child in a given grade in a given context can remain focussed on a pen-paper exercise. For example, from watching the assessment activity closely in May 2014, it was clear that for the Std 4 children an hour for each subject was simply too long. This is a challenge – how to balance the desirable properties of a test with the realities of our children.

School-based versus household based assessments: At the beginning of this exercise, there were discussions about the pros and cons of school based versus household based assessments. The school environment provides a better setting for comprehensive testing of students' skills. Also availability of school based data and participation of key people from the education department makes it more likely that the findings from school assessments will be ploughed into planning and action for the future. However, in the case of this study, the absence from school of almost half of all enrolled children during the assessment makes it difficult to extrapolate the current findings to all enrolled children unless there is evidence that the non-attending children are very similar to the attending children. Constraints of time did not permit surveyors to go find the non-attending children in the village. It should also be noted that at least in the case of reading and math, the estimates from the school based assessment data are very similar to that from household surveys like ASER.

Implications for action

The study focussed on children who had just completed Std 2, Std 4 and Std 6. The assessments were of basic language and math skills. Although this was a cross-sectional study, still it gives clues about what needs to be done to improve basic learning along a continuum of grades from early in the primary school stage to the middle of the middle school stage.

Three clear lessons emerge from the data. Many of these points have been elaborated in the main body of the report and will simply be outlined here:

Basic skills – reading and arithmetic: First, foundational skills like reading and basic math need urgent attention across all grades that were studied. If by the end of Std 2, most children are able to read and understand simple text, then many of the problems that we see today in the higher grades can be avoided. Similarly, if number knowledge till 100 and the ability to do basic operations at least addition and subtraction are in place by the end of Std 2, children can gain math knowledge and skills in subsequent grades quite easily. In the 2014-15 school year, there should be serious efforts to attain these learning goals in Std 2. For the other grades like Std 4 and Std 6, special efforts need to be made to ensure that those who have not attained basic skills do so. Without these skills, these children will not be able to

gain much from continuing to be in school. To achieve such targets, the education system needs to clearly specify learning goals by stages and align all teaching learning activity (such as training, materials and monitoring) systematically to the goals.

Discussion, expression and critical thinking: Although reading is a critical and necessary skill, the data especially for Std 4 and Std 6 shows that it is not a sufficient condition for dealing comfortably with different types of texts. For example, children, who are reading fluently, can do direct fact retrieval tasks from given texts but are unable to do anything else that requires them to go in depth into the content of informative or narrative texts. In particular, tasks that require children to go to different parts of the text, to synthesis meaning, to summarize or to make inferences – all seem to be too difficult to do even with text that is not hard to read. This suggests that in classroom interactions, teachers and children need to spend much more time discussing what has been read and linking that content to what they already know or connect reading material and ideas to everyday life. Typically, much of the time in the classroom in our schools is spent on reading aloud from the textbook and writing on the blackboard (“chalk” time) accompanied by rote learning of textbook material. This kind of teaching practice needs to be transformed to include much more “talk” time with actual discussions and interpretations of what is in the textbook and beyond. Such practices are needed in language classes and even more so in other subjects.

Moving beyond numerical computations to applied thinking and problem solving: One of the interesting facts about Bihar that comes through in most studies of student achievement in the state, is that students perform better in math than in language. This is the case here as well. The data on maths from Std 4 and Std 6 also indicates that while numerical sums are relatively easy to do, the same operations in a word problem form are much more difficult for children to solve. Again this suggests the importance of “talk” and discussion in our classrooms as a way of promoting problem solving skills. It also indicates that all concepts need to be dealt with not just in the traditional numerical form but in a variety of ways that enable children to apply their skills in different contexts. For such practices to take hold in schools, it may be necessary to bring “problem solving” activities into teacher training. It can be achieved by getting teachers to explain problems, encouraging teachers to generate their own word problems, modelling and demonstrations of how this is to be done in high performance schools and by visiting CRCCs and others.

How to deal with a varied range of learning levels in the classroom: The data from May also underlines the existence of wide variations in the abilities of children in the same grade. For example in Std 4 there are a substantial number of children who are Std 1 level and another set of children who are at Std 2 level and less than 20% children who are at grade level. We need to think of how to train, equip and support teachers to simultaneously deal with these multiple groups in the same grade. Alternatively schools need to think about how to reorganize groups across grades to have children at the same level being taught together.

The evidence from this exercise leads us to think of two major changes that are needed in the education system. One has to do with basic skills of children and the other is related to curriculum and expectations. If current curricular expectations have to be met, then teaching-learning activities and conditions have to be reorganized and reworked so that most

teachers can help most children to achieve them. At the same time, it is also worth thinking about whether our curricular standards and textbook content are too unrealistically high.¹

How to handle differences in performance across schools in the same cluster: This study reinforces what is commonly known and experienced – in every cluster there are schools that function relatively well and there are schools that need attention. Looking closely at the cluster report cards we can clearly see these variations cluster by cluster. The challenge for the administration is to figure out ways in which we can productively use these variations to improve school functioning and student performance of all schools. Immediate steps could include:

- Detailed discussion and dissemination of cluster report cards with the respective Cluster Coordinators as well as similar discussions with DIETs and all Cluster Coordinators in a district. If the two sampled clusters are representative of clusters in the state, what kinds of actions need to be taken to improve current status?
- Evidence based movement plans can be drawn up by Cluster Coordinators for monitoring schools in the cluster. The CRCC could spend more time in the relatively weaker performing schools. Actual instruction by CRCCs, modelling of how lessons can be taught, special training of teachers, reorganization of groups for teaching, increase in facilities/inputs directly needed for teaching-learning – are all things that can be tried.
- The better performing schools and good teachers can be used as “models” for others. This can be done by holding meetings or “guru goshtis” in rotation in the well performing schools so that actual classes can be seen by others or by taking such teachers from time to time to other schools to demonstrate how they organize teaching and how they carry out instruction. Pairing of “good” schools with “weaker” schools can also be a strategy.

Conclusion

In conclusion, large scale assessments of student achievement provide opportunities for doing a lot more than data collection. Taking advantage of these opportunities, it is essential that we think about how this can lead to building capacities for assessment and instruction, for trying new and more appropriate measurement methods, discussing learnings from the process, understanding findings and connecting them to the next stage of planning and implementation.

¹ See paper by Pritchett and Beatty (2012) on the negative consequences of overambitious curriculum.