Azim Premji Foundation

School Quality Assessment:

Multiple Matrix Model as a Viable Alternative For Large Scale Assessment of Learning Achievement

June 2009

D D Karopady and S Giridhar
Azim Premji Foundation, Bangalore

Azim Premji Foundation

Table of Contents

Summary

This paper is based on the experiences and results of school quality assessment carried out on a large scale carried out by the Azim Premji Foundation under its Learning Guarantee Program (LGP). This program is now being implemented in a few other states while the Karnataka Government has launched a similar effort to assess school quality through the Karnataka School Quality Assessment Organisation (KSQAO). The LGP model of school assessment may have some advantages but at the same time these methods are time, human resource and effort intensive. It involves testing every child in all the classes for each competency and written testing is accompanied by oral testing. After a couple of years of administering these tests it was clear that it is necessary to evaluate alternative models that are significantly superior from the point of view of scalability and sustainability in the system while at the same time provide comparable results. In this paper, we begin with the premise that there is a place for large scale assessment that can be of significant benefit to all stake holders.

Sampling is an option for making the process more efficient. However, any sampling implies that the results of the sample will be at some variance with the results of the total. There are several ways in which the sampling can be carried out. In an attempt to generate alternative models, the two specific models have been explored:

Assessment of all children in the class but in only some of the classes instead of all four (i.e. one of the four classes or a combination of two classes).

Assessment of all classes using a multiple-matrix design of tests (i.e. every child every class but in sample of competencies)

These two methods have been simulated using the database of test scores of every child (261590 children in all) in classes 1 to 4 in Language and Math, in the 1868 schools evaluated in Learning Guarantee Program in Karnataka in 2005. In order to decide if the alternative sampling methods are acceptable or not, the results generated for these two models are compared with the original results of each school.

The data based analysis clearly shows that although simple sampling (involving all children but only in some classes) is quicker and cheaper, the results are at substantial variance from the results using the traditional (every child every competency assessment) method. For example, if only one class is selected at random and used to represent the entire school, the average of over 60% schools varies from the actual by over 10%. If two classes are used, the difference is over 10% in almost one third of schools. This is still unacceptably large.

On the other hand the Multiple Matrix method (i.e. every child, every class but in sample of competencies) is not only significantly more efficient (needing just about 35% to 40% of resources) but the results are extremely close to the traditional assessment (within + or - 3% in case of most schools). In this design, instead of different tests for different subjects, all the subjects are combined to form only one test. Further, this one test is split into three matched question papers each with one third the number of questions. Thus, each child is tested for only some competencies reducing the load on the child. The three question papers are administered to three random groups of children in the class. The performance of the 'class' is worked out based on average of all three question papers.

After the simulated results indicated that the Multiple Matrix Method was a better option, Azim Premji Foundation tried this method in Rajasthan Learning Guarantee Program with very good and reliable results. The Foundation recommended this option to Karnataka School Quality Assessment Organisation (KSQAO) who implemented this in the 2009 assessment. It is hoped that those state education functionaries who are interested in such assessment will also consider the suitability of employing the Multiple Matrix Method in the light of these findings.

1. Introduction

This paper is based on the experiences and results of school quality assessment through learning achievement tests on a large as well as geographically dispersed scale. The Learning Guarantee Programme[1] ran as a pilot for three years between 2002 and 2005 in 1868 primary and higher primary schools of North East Karnataka. The Government of Karnataka then expanded the experiment to 6464 schools in the 202 educational blocks of the state in 2005. This subsequently metamorphosed into the Karnataka School Quality Assessment Organisation which carried out its first assessment in January 2006. Some other states have also embarked on large scale school quality assessment or school grading on the basis of various indicators, including learning achievements of children.

It is not within the scope of this paper to debate the need, the merits or the downsides of large scale testing to assess school quality. The authors begin with the premise that there is a place for assessment and if done right it has enormous value for: the teacher in the class room; teacher educators and the system. If the stakes and stress are kept low, if the results are consciously used to provide systemic inputs to improve quality and if schools delivering quality are recognized such assessments can be very meaningful.

The Learning Guarantee Programme model of school assessment may have some advantages but at the same time these methods are time, human resource and effort intensive. It is perhaps necessary to identify all the areas where it would be worthwhile to consider alternative models that are significantly superior from the point of view of scalability and sustainability in the system while at the same time providing comparable results.

The authors have at their disposal the child wise, item wise data from the evaluation of 1868 government primary schools for every child in classes 1, 2, 3 and 4; their learning achievement scores on competency based oral and written tests in Language and Math. This has been used to run a simulation of alternative models based on sampling designs and test them for reliability, accuracy. The purpose of assessing various models is to identify the alternative that would best achieve a significant reduction in the time, human resources and effort compared to the method of assessing every child in every class for all the competencies while at the same time provide results that are very close to the original method.

With many state education departments in the country already implementing or wanting to implement low stakes, large scale school grading initiatives based on child friendly competency based tests (and not testing rote learning), the importance of a cost effective model that meets this limited purpose gains immediate significance. By opting for the most resource effective model, the state would save a lot of time, energy and funds that could instead be channelized for the academic support functions that need to logically go hand in hand with assessment.

---

[1] Leatrning Guarantee Programme is a joint initiative of the state government and Azim Premji Foundation. The programme ran as a pilot in 7 districts of Norh East Karnataka between 2002 and 2005. A key element of the program is that it sought voluntary participation of schools. 1868 schools volunteered and were assessed on the criteria of enrolment, attendance and learning achievements of the children in their primary grades.

LGP has two key objectives: One objective is to create a spirit of accountability among schools and education functionaries for the learning of every child. The other key objective of the programme is to advocate a classroom teaching learning reforms through systemic shift in assessment - from the traditional test of rote learning to test of a child's understanding, application and problem solving ability.

In the pilot program in Karnataka, every child in classes 2, 3, 4 and 5 was tested through oral and written tests for competencies of their previous class. The tests were not based on the text book but were designed specially to evaluate learning, understanding and application of the expected competencies of the children. The evaluation of schools in the Pilot programme was done in July September of each year and hence the children were assessed for the previous grade's competencies. Oral tests were administered for about 15 minutes per child for each subject, while children were provided 90 minutes for written tests for each subject. A team of 4 independent evaluators conducted these tests. A school that has about 80 children in Classes 1 to 4 (or 2 to 5) would take 4 members of an evaluation team, 3 days to complete the evaluation. During this evaluation, the team also collected data pertaining to enrolment and attendance of every child.

1. The Learning Guarantee Programme Model of School Assessment

Alternative Models of school assessment is the subject of this paper. This section describes the essential features as well as the operational details of the Learning Guarantee Program model of assessment of schools to set the context. A more detailed description of the Programme is available in various reports and papers[2] .

The articulation of the concept[3] of Learning Guarantee Programme has graduated over a period of time starting from 2002 when it was first launched in North East Karnataka. However we give below the criteria used in the Karnataka Learning Guarantee Program for the schools to qualify

Enrolment	100% of children in the 6 – 14 age group in the habitation must be enrolled in school
Attendance	Minimum of 90% of the enrolled children must have attended at least 75% of total number of working days in school
Learning	Minimum of 60% of all children enrolled in classes 1 -5 score 90% on competency based tests.

These criteria meant that evaluation would require collection of data on learning achievement of every child in the appropriate classes besides data on enrolment and attendance.

The evaluation of learning achievement of every child can be extremely time and effort intensive task especially since in the lower classes the children are best assessed using individual oral questions. In addition, since the assessment was by an 'independent' team, the children would be unfamiliar with the team members and the processes and hence feel uncomfortable. There was thus a need to address this aspect in the process design.

The evaluation process followed for the program comprised the following key steps:

A. Collecting data on enrolment of children from the habitation
b. Collecting data on the attendance details of every child for the appropriate time period
c. Spending time with the children to make them feel comfortable and minimise the 'fear' of assessment to the extent possible
d. Assessing each child for written competencies as well as oral competencies in the appropriate subjects
e. Correcting the answer sheets and checking the papers

The evaluation of schools in the Pilot programme in Karnataka was done in July September and hence the children of classes 2, 3, 4 and 5 were assessed for their respective previous grade's competencies in Language and Mathematics through written tests of 90 minutes and oral tests of 15 minutes per child per subject. A school that has about 80 children in Classes 1 to 4 would take 4 members of an evaluation team at least 3 days to complete the evaluation process.

---

[2] "The Learning Guarantee Programme", Seminar 536 April 2004
" Learning Guarantee Programme: 2003", Learning Curve, Issue II, March 2004
"The Learning Guarantee Programme: A Learning Journey 2002-05", D D Karopady and S. Giridhar, UNESCO Page 194-214, Pratham UNESCO New Delhi, publication: City Children, City Schools, October 2005
" Assessment Reforms Through Voluntary Participation of Schools" , D D Karopady, S. Giridhar, and Umashanker Periodi at National Conference on Pupil Assessment System at Elementary Level, NCERT March 06
" Case Study of LGP Among 15 Promising Practices in the Country" ILO, Edited by Vimala Ramachandran, June 06
" Process Manual for Evaluation of Schools in Learning Guarantee Programme", Version 2, June 2004: Dileep Ranjekar, S. Giridhar (Azim Premji Foundation) and Prof. Nayanatara (Indian Institute of Management, Bangalore)

[3] Concept Note: "Learning Guarantee Program - An Attempt to Create Examination based Classroom Reforms", Version 4, May 2006

A typical 3 day schedule for the team would be as follows:

Day	Task	Elements of the Task
1	Enrolment and attendance data Verification	The 4 members verify and tally the enrolment data and the data of the children's attendance from the registers. Schools are provided specific formats in which they fill the data and submit to the evaluators. This task takes about 90 minutes.
	Writing the details on the question papers	For every child in the Attendance register, the evaluators write down child wise details (school code, child's name, sex etc) in the space provided on the question - answer paper. This takes about an hour
	Written and oral tests for four classes – subject 1 - (Language)	Each member takes responsibility for one class and all four evaluators simultaneously conduct the written tests for the children of their class. The written test duration is 90 minutes. Each evaluator then conducts oral test for the children in the respective class for Language. Each oral test takes about 15 minutes per child. Thus in about 3 hours an evaluator can conduct 15 oral tests. Evaluators check and mark every answer paper at the end of the same day.
2	Written and oral tests for four classes – subject 2 - (Math)	Each member takes responsibility for one class and all four evaluators simultaneously conduct the written tests for the children of their class. The written test duration is 90 minutes. Each evaluator then conducts oral test for the children in the respective class for Math. Each oral test takes an average of 15 minutes per child. Thus in about 3 hours an evaluator can conduct 15 oral tests. Evaluators check and mark every answer paper at the end of the same day. In case there is time left on the second day, after completing these tasks, evaluators complete the oral tests of children left over from Day One (Language). Evaluators who finish their class oral tests move to the class where oral tests have not been completed to help their team member complete this task.
3 +	Completion of Oral Tests	It is not possible to conduct more than 20 oral tests in a day for one evaluator on the first three days because written tests are conducted in the forenoon. Therefore in schools with more than 20 children per class, evaluators need to stay beyond three days to complete the oral tests. Thus for a school whose strength is between 80 and 180 children in 4 classes (estimate of maximum 45 children / class), Day 4 is required. And for every additional 100 children a further day would be required.

Thus, as can be seen, the evaluation process typically took a team of four about 3-4 days (or even more) to complete. If a third subject - Environment Science (EVS) were to be added, the time required would go up even more. It must be noted here that this detailed evaluation was preceded by careful selection and recruitment of the team members and a detailed 4 day residential training program on the processes to be followed. In Learning Guarantee Program Karnataka, 460 evaluators in 115 teams of 4 each were engaged for three months (74 working days July to September) to conduct evaluation of 1868 schools.

Since every child was being assessed, the question papers had to be printed in such quantities. In Learning Guarantee Program Karnataka, for the evaluation of 1868 schools, 261590 question papers in Math and Language each were printed; packed school wise in the required quantities and dispatched to the block head quarters. At the block head quarters, one person was assigned the task of converting the child wise score sheet for the school into an "Intelligent Character Reader sheet" (some use the "OMR" sheet) for processing by computer. For transcribing the answer sheets of over 250000 children, it took 120 persons about two weeks. Monitoring, supervision and technical support to these evaluators and transcribers is critical to maintain quality, integrity and morale. One supervisor guided and monitored three teams.

We have not avoided mundane operational details while describing the process in order to showcase to the readers how intense this model is in terms of logistics, time, effort and human resources.[4]

[4] For a report on L G P school evaluation process, "Learning Guarantee Programme Uttaranchal: School evaluation process in Uttarkashi and Udhamsingh Nagar", Naitra Muralykrishnan, May 2006

As would be obvious, this kind of assessment is also cost intensive. A bird's eye view of costs is provided with an example below.

1. 10 Districts: 100 Educational Blocks: 1000 clusters
2. Approximately 18000 primary and upper primary government schools
3. Estimated 15 lac children in classes 1 to 4 in these 18000 schools
4. Evaluation to be done in 1 month
5. 3 days / school for 4 member team: Average process schedule

	Item	Estimated Cost
1	Question Papers: @ Rs. 3 per child for 15 lac children	Rs. 45 lacs
2	ICR sheets to transcribe children's scores: @ Re. 1 per child	Rs. 15 lacs
3	Scanning and data processing: @Rs 1.85 per child	Rs. 27.8 lacs
4	School wise performance report feedback; @ Rs. 60 per school	Rs. 10.8 lacs
5	Transportation of material: @ Rs. 100 per school	Rs. 18 lacs
6	Honorarium or TA for 10000 evaluators: @ Rs. 70 per day for 24 days	Rs. 154 lacs
7	Communication material about the program, sample question bank etc. @ Rs. 40 / school	Rs. 7.2 lacs
8	4 day training for 10000 evaluators @ Rs. 120 per evaluator	Rs. 56 lacs
9	Orientation, follow up meetings, development of assessment tools in 10 districts: 6 workshops during the year	Rs. 2.5 lacs
10	Honorarium to external experts for capacity building exercises	Rs. 1.5 lacs
11	Reward and recognition program for schools, teachers; assumption @ Rs. 20000 per block	Rs. 20 lacs
12	Follow up meetings at district, block and cluster to discuss the performance feedback, analysis and develop an improvement plan	Rs. 12 lacs
	TOTAL	Rs. 370 lacs

Costs: Rs. 2050 per school. Human Resources required: 10000 persons for one month; i.e. 100 well qualified, motivated, dedicated volunteers in every block or taluk.

Later when KSQAO conducted school quality assessment, they followed a similar process but with some modifications - they decided to assess children in classes 2, 5 and 7 to represent the school (class 1 to 8). and they did away with oral tests for class 5 and 7. Despite this, the logistics was like a gigantic military operation:

- 26 lac children in around 40000 schools.
- Question papers printing and individual school wise bundles at 202 blocks for 40000 schools.
- 25000 evaluators to conduct the assessment 150 master trainers
- 26 days in January 2006 to complete this exercise.
- Block and cluster resource persons exclusively assigned to monitor and support the evaluation.
- A quality control and scanning of all child score sheets at every District (DIET office)
- Results declared by June 2006 4 months after the evaluation.

It is in this context that we raised fundamental questions:

If the objective is to assess school quality why is every child being assessed?

Why is it necessary to assess learning outcomes of every child in every subject and virtually every learning area / competency?

What are the other ways and means of arriving at the same conclusions (school performance against the Learning Guarantee Programme criteria) but which can be done at less than half the cost, half the time, half the human resources employed in this model?

1. Exploring Alternative Models for School Quality Evaluation

We have seen in the previous section, that the method of assessing every child in every class for every competency requires a commitment of around Rs 2050 per school and nearly 10000 evaluators for a month in order to evaluate 18000 schools. This is a huge commitment of resources; the processes, logistics are stressful, and the duration causes high degree of fatigue. All these will ultimately affect the overall quality of the evaluation and even if the state keeps on doing this it will soon reduce to a humdrum routine. Thus, there is a need to come up with alternative models that require much less money, human resources and time, and at the same time yield results that are as robust, reliable and accurate as the existing model.

Sampling is a possible option for making the process more efficient. However, any sampling implies that the results of the sample will be at some variance with the results of the total. There are several ways in which the sampling can be carried out. In an attempt to generate alternative models, the authors have explored specific models. The underlying principles for sampling are: a) for assessing school quality, we need not assess every child; b) it may not be necessary to assess every class in a school and c) since these tests are not replacing any formal annual examination, it is not necessary to test every child for every learning area/ competency.

The sampling options considered are:

I. Assessment of all children in the class but in only some of the classes instead of all four (i.e. one of the four classes or a combination of two classes).
II. Assessment of all classes using a multiple-matrix design of tests (i.e. every child every class but in sample of competencies)

These two methods have been simulated using the database of test scores from the 1868 schools evaluated in Learning Guarantee Program Karnataka 2005. This database comprises the item wise score of every child (261590 children in all) in classes 1 to 4 in Language and Math, in these 1868 schools. In order to decide if the alternative sampling methods are acceptable or not, the results generated for these two models are compared with the original results of each school for the following parameters:

a. the average performance of the class (average percentage of 'all correct' responses)
b. the number of children above the 'cut-off' level in this case 90% achievement level
c. winning status of a school on achievement parameter

I. ALTERNATIVE MODEL 1 - Assessment of all children in the class but in only some of the classes instead of all four (i.e. one of the four classes or a combination of two classes)

This model proposes that it will be adequate to test randomly any one or combination of two classes in a school to obtain an accurate and reliable evaluation of the school's performance under the criteria of “Learning achievement”. Can a few classes be used to represent the entire school? This is what the results of the simulation show:

The average achievement levels of different classes are compared with the average achievement levels of the school (all 4 classes put together).

Achievement (Language + Math) in %

Class 1	Class 2	Class 3	Class 4	Tot – 1 to 4
61.6	61.9	55.8	59.1	59.5

If all the 1868 schools are considered together, the average performance in any class appears to be about 3 percentage points (or about 5%) of the total average. For example the average achievement of class 1 at 61.6% is just 2.1 percentage points (or 3.5%) higher than the total school's average achievement at 59.5%.

However, if this analysis is disaggregated and carried out at a school level, the differences emerge more sharply.

Number of schools with difference in excess of 10%

Only class 1	Only class 2	Only class 3	Only class 4
1080 (57.9%)	1012 (54.2%)	1019 (54.7%)	1050 (56.6%)

In this case, a large majority (close to 60%) of the 1868 schools have difference in average achievement of over 10% as compared to total school average. For example, if only class 3 average were to be considered, in case of 1019 schools (54.6% of the 1868 schools), the total school average performance varies by more than 10%. The situation is similar for any other class considered separately as well. Thus just a sample of any one class among the 4 lower primary classes does not serve the purpose of being a close approximation of the entire school average.

The scenario improves to some extent (as can be expected) if two classes are taken together.

Number of schools with difference in excess of 10%

Classes 1+2	Classes 1+3	Classes 1+4	Classes 2+3	Classes 2+4	Classes 3+4
723 (38.7%)	628 (33.6%)	571 (30.6%)	564 (30.5%)	594 (31.8)	711 (38.1%)

In this situation, between 30% and 40% schools have difference over 10%, a much better proposition than that close to 60% noticed in the case of only one class. The class combinations which are the closest (relatively) are classes 2+3 and classes 1+4. But even in these two selections at 30% of the schools (564 in the case of class 2+3 and 571 in the case of class 1+4) show a difference of more than 10% when the average cumulative scores of the two classes are compared with average cumulative scores of all 4 classes.

To use another basis of analysis proportion of children achieving 90% or more in the assessment a similar pattern is discernible. In the total sample, when all the children in all the classes were assessed, 152 schools would have emerged as winners (on the LGP criteria). However, if combinations of classes were to be used instead of all classes, some schools would suffer (or benefit) unfairly as can be seen below.

Number of winning schools

All classes	Classes 1+2	Classes 1+3	Classes 1+4	Classes 2+3	Classes 2+4	Classes 3+4
152	164	142	161	148	173	155

Here too, a combination of classes 2 and 3 or classes 3 and 4 appears to come closest to 'all classes' option.

These differences are too large to be comfortable. Thus, the options of considering only one or some two of the classes out of the four classes do not appear to be the correct approach. Certainly, taking three of the four classes would appear better than this. However, on the ground, that would hardly offer any significant efficiency in terms of saving time, effort or resources as compared to an assessment of all four classes. The objective of this

exercise is to find the model that provides the quality, reliability and accuracy as the assessment of all four classes but by using significantly less time, money and human resources. Hence, combination of classes as an alternative to the whole school does not seem an acceptable option.

II. ALTERNATIVE MODEL 2 - Assessment of all classes using a multiple-matrix design of tests (i.e. every child in every class but only in sample of competencies)

This model is based on the premise that in order to evaluate a school's performance against its commitment to ensure every child is learning it is not necessary to test every child for every competency commensurate to her age or class. This model explores the possibility of assessing children on a few randomly selected competencies relevant to their age and class. The design is such that every competency is equally represented in the test by ensuring distribution of competencies across the children in a class. Thus if we have 20 items in a Math question paper we distribute randomly some questions to 1/3rd of the children, some to the next 1/3rd and the remaining items to the balance 1/3rd children. Ralph Tyler and John Tukey, the Founders of National Assessment of Educational Progress selected the multiple matrix sampling to survey the knowledge of students across the nation with respect to a broad range of content and skills, and to report the relationships between knowledge and a large number of educational and demographic background variables. The idea for the model proposed here is derived from the Multiple Matrix sampling method.

In this design, instead of two (or three) different tests for different subjects, all the subjects are combined to form only one test. However, to reduce the load of questions on the children, the combined test is split into three 'matched' sets with questions randomly assigned to each set. The three sets are then administered to three random groups of children in the class. The class average is worked out based on the average of all three groups (and hence all three sets) together. This obviously offers significant efficiency on the ground (reducing the time, effort and also the cost) to almost a third of what would be required if all tests were separately administered to all children.

However, how representative would this be? How close would the average achievement levels be in the two methods?

It is first necessary to establish as to how close the performance (based on average achievement) between a full scale assessment and a multiple matrix mode is using three sets of matched questions. Once again the authors used the data base of the 1868 schools to simulate the experiment and the results thus obtained were compared with the Learning Guarantee Programme 2005 results.

The data of achievement of the 1868 schools was split into three data sets with randomly assigned children and randomly assigned questions to the three sets. The average achievement for each class has been worked out using average of the three matched sets and compared with average of the full scale separate tests as earlier. The comparison at an overall level is as follows

Average achievement levels

	Class 1	Class 2	Class 3	Class 4
Avg. for full scale test in all subjects	61.62	61.87	55.80	59.06
Avg. for multiple matrix test	61.59	61.89	55.79	59.07

This table shows how close the averages for all the schools together appear in the two methods.
If the averages are separately worked out at the school level, the numbers look like follows:

Number of schools

	Class 1	Class 2	Class 3	Class 4
Total number of schools	1864	1866	1864	1854
Difference between 5 and 10%	82	142	120	105
Difference between 2 and 5%	413	508	480	428
Difference less than +/- 2%	1352 (72.5%)	1174 (62.9%)	1213 (65.1%)	1286 (69.4%)

(The difference is measured in percentage of the average if the three full subject tests were administered separately)

The table very clearly shows that the average performance using the Multiple Matrix Model is very close to the full test average and the differences are very small. For example, in case of class 1 average, in as many as 1352 schools (72.5% of schools), the difference between Multiple Matrix calculation and Full Scale test is less than 2%. Further, using the number of children who score 90% or more as the basis of analysis also shows that the numbers are identical in both methods.

It must be noted here that if we look at small schools (class size less than 10) the error margin increases. This is a limitation of this method. The table below highlights this aspect.

A Break up of Schools with Difference less than +/- 2%

	Class 1	Class 2	Class 3	Class 4
Class size below 10	157 (50.5%)	109 (37.3%)	126 (43.0%)	131 (50.8%)
Class size 10 and above	1195 (76.9%)	1065 (67.7%)	1087 (69.2%)	1155 (72.4%)

Analysis of oral and written achievement correlation

The achievement data was also analyzed to determine the correlation if any, between the performance of the children in the oral section and the written section.

There would certainly be differences in the performance in oral and written assessment at the individual child level. The attempt here is to look at the relationship between the two at the overall school level. The average performance of the children in oral and written assessment was worked out separately for each class in each school. The average in oral test was treated as the 'dependent variable' (y) and the average in the written test was treated as the 'independent variable' (x) in a regression analysis. A correlation analysis was also carried out. This was done for each class separately. The findings are summarized below.

Regression analysis

	Class 1	Class 2	Class 3	Class 4
Multiple R	0.8231	0.8091	0.8036	0.8296
R Square	0.6776	0.6547	0.6458	0.6883
Adjusted R Sqr	0.6774	0.6545	0.6456	0.6881
Standard Error	11.0651	10.5564	12.2159	8.8689
Observations	1864	1860	1860	1860

ANOVA

F	3912.79307	3522.23744	3387.53646	4103.1569
Significance F	0	0	0	0

Correlation

Oral-Written avg	0.82292	0.80911	0.80361	0.82965

The strong correlation between oral and written achievement levels is unmistakable[5]. Admittedly, certain competencies and skills can be assessed only with oral testing. However, if an overall assessment of a school or class is desired, this analysis suggests that a written measurement process serves the purpose adequately.

All the above analysis shows that it is preferable to use a multiple matrix design of assessment to bring about operational efficiencies.

It is important to note two critical aspects of the model - (a) the design of the tools (question papers) and (b) the implementation (evaluation process). Both these need to be addressed carefully.

Tool design (Question paper design)

The tool has to be designed such that the different sets of question papers have an equal and even distribution of competencies, levels of ease or difficulty and are equally challenging. This is not just a task of allocating the questions randomly to the three sets of papers. Considerable thought, preparation and pre testing of the items would be needed for this.

Implementation (Evaluation process)

It is not enough for the question paper sets to be designed as equally challenging. The sets need to be equally distributed across all the children who are evaluated to ensure that all the relevant competencies are evaluated for a given class. It may be tempting to pre decide which child gets which question paper set based on the enrollment number. This however could lead to problems of imbalance in completed sets of papers in the event of absence of children allocated to a certain set. Hence, the distribution has to take place based on the actual number of children present at the time of the test.

Also, the issue of 'reading out questions' to the children would need to be looked at seriously. This is important for children typically in lower classes (class 1 3 or even up to class 5). The presence of different sets of papers in the same class poses challenges. This needs to be thought through. Of course, if oral questions (or reading out of questions) are done away with, then this is a non issue.

[5] Another analysis done elsewhere in the Andhra Pradesh Randomized Evaluation Study also showed that the correlation between oral and written scores is over 0.75 for classes 1 to 5.

1. Multiple Matrix Model experiment in Rajasthan

The Learning Guarantee Program in Rajasthan provided Azim Premji Foundation the opportunity to experiment with the use of the Multiple Matrix Model of assessment in both the baseline (November 2006) and in the school evaluation conducted in 1039 schools of Tonk and Sirohi Districts that participated in the program. This provided Azim Premji Foundation on the ground experience with the model and enabled it to document the benefits as well as downsides of using this model.

In Rajasthan, the competencies for all the subjects of each class were divided across three sets of question papers (Set A, B & C). The three sets were equally distributed to the children present on the day of evaluation. Thus in about 60-90 minutes, every child in a class was assessed in all the subjects in a manner that 1/3rd of the children present answered Question Paper set A, another 1/3rd answered Set B while the remaining 1/3rd answered Set C. This enabled the completion of evaluation in a short time with significantly less resources compared to the traditional model. The Rajasthan Learning Guarantee Program team reported that:

i. On an average, a school was evaluated by four evaluators in 1 day as against 3 days in the traditional model
ii. A child was able to complete his/her assessment in all the subjects in a time span of 60 to 90 minutes, whereas in the traditional model it was spread across three days of evaluation
iii. There was sufficient time during the rest of the day for the evaluator to conduct oral tests.

Analysis of the results (tests conducted in November 2007) show that the performance of children across the three sets of question papers is quite similar.[6] The difficulty level across the three sets when viewed at an overall level is very similar but there are variations within each district. The table below shows the average learning levels displayed by children on the different sets of question papers in Sirohi and Tonk Districts

	Set A Avg	Set B Avg	Set C Avg
District - Tonk	62.0	49.5	62.5
District - Sirohi	52.4	59.8	51.0
Total	57.1	54.7	56.8

Also, the three test sets were quite evenly distributed as evident from the table below.

No of children	Set A	Set B	Set C
District-Tonk	9750	8824	7873
District-Sirohi	7618	7077	6629

Another indicator of a good question paper and good evaluation processes is the minimum difference in performances on Oral and Written questions. It is expected that children will perform better in oral than in written; however, if the averages are lopsided, it indicates that subjectivity may have played a strong role during the evaluation processes. The difference between oral and written averages in both the Districts was not too large (less than 8%) in this case.

The results from the administration of the Multiple Matrix Model in Rajasthan are thus very encouraging - a model that operates at less than 1/3rd the cost (both time and resources) of the models of evaluation that have been so far followed in large scale evaluations such as the Learning Guarantee Program or KSQAO.

[6] Azim Premji Foundation has partnered with Vidya Bhawan Society Udaipur for the Learning Guarantee Program in Rajasthan. The academic resource unit of Vidya Bhawan led by Dr. Hridaykant Dewan took an active role in conducting assessment tool development workshops and finalizing the question papers

5. Savings on Resources using Multiple Matrix Model

Taking the same illustrative example that we described in earlier section, the resource utilisation has been compared for the two methods:

1. 10 Districts: 100 Educational Blocks: 1000 clusters
2. Approximately 18000 primary and upper primary government schools.
3. Estimated 15 lac children in classes 1 to 4 in these 18000 schools.
4. Evaluation to be done in 1 month (22 working days)

Sr No	Item	Existing LGP or Pre 2008 KSQAO Method	Multiple Matrix Method
1	Question Papers: @ Rs. 3 per child for 15 lac children	Rs. 45 lacs	Rs. 15 lacs
2	ICR or OMR sheets to transcribe children's scores: @ Re. 1 per child	Rs. 15 lacs	Rs. 15 lacs
3	Scanning and data processing: @Rs 1.85 per child	Rs. 27.75 lacs	Rs. 27.75 lacs
4	School wise performance report feedback; @ Rs. 60 per school	Rs. 10.8 lacs	Rs. 10.8 lacs
5	Number of days to evaluate one school with 4 members in a team	3 days	1 days
6	Number of evaluators required to complete the evaluation in 1 month	10000	3300
5	Transportation of material: @ Rs. 100 per school	Rs. 18 lacs	Rs. 18 lacs
6	Honorarium or TA for evaluators: @ Rs. 70 per day for 22 days	Rs. 154 lacs	Rs. 51 lacs
7	Communication material: About the program, sample question bank etc. @ Rs. 40 per school	Rs. 7.2 lacs	Rs. 7.2 lacs
8	4 day training for evaluators @ Rs. 120 per evaluator	Rs. 56 lacs	Rs. 19 lacs
9	Orientation, follow up meetings, development of assessment tools in 10 districts: 6 workshops during the year	Rs. 2.5 lacs	Rs. 2.5 lacs
10	Honorarium to external experts for capacity building exercises	Rs. 1.5 lacs	Rs. 1.5 lacs
11	Reward and recognition program for schools, teachers; assumption @ Rs. 20000 per block	Rs. 20 lacs	Rs. 20 lacs
12	Meetings at district, block and cluster to discuss performance and develop improvement plan	Rs. 12 lacs	Rs. 12 lacs
	TOTAL	Rs. 370 lacs	Rs. 198 lacs
	Cost per school	Rs. 2050	Rs. 1100
	Human Resources	10000 persons	3300 persons

The above table shows that the Multiple Matrix Model can help save 43% of costs as compared to the traditional method. Similarly, the saving on human resources is huge as only 1/3rd the human resources are required.

6. Conclusions

Large scale assessments in the traditional approach require huge effort, time, human resources and of course cost. There are a few alternatives available to make the process significantly easier and cost effective while ensuring comparable results. The analysis above clearly shows that the Multiple Matrix Model approach is significantly more efficient without compromising much on the reliability and validity of the data. The costs can be easily slashed by about 50% while the human resource requirements can be reduced by as much as 65% as compared to the traditional approach.

Admittedly, three aspects need to be borne in mind

a. The question paper needs to be designed with care to ensure that the versions are of similar difficulty level and test similar number of competencies
b. The implementation of the test has to ensure that all the versions of the test are answered by approximately the same number of children in any class
c. Small class size (less than 10 children) can introduce a greater error in the estimation

This kind of assessment and analysis is suitable when the class as a group and the school as a whole are being measured. If however a detailed, child wise understanding of attainment of competencies and skills were needed and if feedback is required to be provided for each child then there is no alternative to testing every child for each competency and skill.

The authors hope that the state education functionaries will now examine the suitability of the Multiple Matrix Model for their purpose of assessing learning achievement for classes and schools in the light of these findings.