Survey Design and Precision of ASER Estimates

Bharat Ramaswami & Wilima Wadhwa[1]

1. Introduction

The purpose of ASER (Annual Status of Education Report) is to obtain reliable estimates of the status of children's schooling and basic learning (reading and arithmetic level) in rural India. The survey is undertaken in rural India every year and there have been four such surveys starting from 2005. Every year a core set of questions regarding schooling status and basic learning levels are posed to children in the 3-16 age group.

The objective of this study is twofold. First, we examine whether the sampling methods used in ASER generate estimates within a reasonable margin of error. This is done at the state as well as the district level. In this second part of this study, we examine the issue of sample design. In particular, we seek to answer the question that if greater precision of district estimates is desired, then how should the sample design be modified?

2. Description of Survey

The survey is undertaken in all rural districts of India. The survey is designed to be a household survey. Within each district, 30 villages are randomly chosen[2] and in each village 20 households are randomly picked for a total of 600 households per district. ASER reports district and state level estimates. The 2007 survey samples 564 districts out of a total of 584, 15,958 villages, 317,116 households and 715,620 children in the 3 – 16 year age group.[3]

The key feature of ASER is that it is a rapid assessment survey. As a result, the survey instrument is short and its focus is on the assessment of basic learning. Since 2006, ASER has included younger children in the sample. However, children aged three and four are not tested. All that they are asked is whether they attend any kind of pre-school (such as anganwadi). Older children (aged 5 to 16) are queried about school enrollment and were tested in basic reading and arithmetic. These tests have been carried through all the ASER surveys. Additional tests (such as comprehension, English, currency tasks) have been administered in some years but not in all years.

All children (irrespective of age) are administered the same tests. However, the tests have different levels graded by difficulty. The reading test starts off with an easy

[1] Financial support from Hewlett Foundation is gratefully acknowledged. We thank Dana Schmidt for insightful comments as well as Arka Roychowdhuri and Swagata Bhattacharya for excellent research assistance.

[2] Villages are chosen from the 2001 Census Directory using PPS (Probability Proportional to Size).

[3] These numbers are based on the ASER 2007 data.

paragraph (coded as Level 1 text[4]). If the child reads the paragraph fluently, then the child is asked to negotiate a longer text at a higher level of difficulty (Level 2 text). On the other hand, a child who has problems with Level 1 text is moved on to a simpler task – the reading of words and if this is also too difficult, the child is tested on recognition of letters.

The arithmetic test is similarly constructed. The initial task is a subtraction problem.[5] A successful child is then tested with a division problem.[6] If subtraction is too hard for the child, the child is tested on recognition of two digit numbers (11 – 99) and if even that it too hard, the child is tested on recognition of recognition of one digit numbers (1 – 9).

3. Variables

We picked the following learning outcome variables that are presented in the ASER report.

- % children in age group 3-4 years who are in Anganwadi or other pre-school.
- % children in age group 6-14 years who are out of school
- % children in age group 6-14 years who are in private school
- % children in class 1-2 who can read letters, words or more in own language
- % children in class 1-2 who can recognize numbers (1-9) or more
- % children in class 3-5 who can read level 1 (Std 1) text or more in own language
- % children in class 3-5 who can subtract or do more

All these are reported at both the district and state level. We chose the following states: Assam, West Bengal, Himachal Pradesh, Bihar, Rajasthan, Andhra Pradesh and Karnataka. The choice was made to obtain a fair representation of states from different regions of the country.

4. Statistical Issues

The statistical precision of district level estimates is an issue because of the ASER sample design – namely clustering and absence of stratification at the village level.

In a design without clustering, children in the relevant age group would be directly sampled. Not only is this expensive (in terms of survey time), it is difficult to have a reliable population frame that could be used for sampling. Instead ASER employs a two-stage clustering design. The first stage clustering happens when villages are randomly picked. The second stage clustering is when households within a village are randomly picked and the children belonging to that household are tested.

---

[4] Level 1 text is at the level of Class 1 and Level 2 text is roughly a Class 2 level text.
[5] Two digit from two digit with carry-over.
[6] Three digit by one digit with remainder.

While this is an inexpensive and practical way of sampling children, it is well known that clustering increases the variability of estimates. The increase in variability (relative to sampling without clustering) due to clustering is called the design effect. This issue is relevant to the analysis of statistical precision of estimates at the district level.

One way of increasing precision at the district level would have been to stratify the village sample according to age of children or parental background (wealth, mother's education). However, this would require a prior household listing and this is expensive in terms of time and resources.

The ASER sample is stratified, however, at the district level. In so far as outcomes within a district are more homogenous than across districts, stratification within the district would lead to more precise estimates at the state level.

5. Estimates of Precision

Precision of sample estimates is calculated taking into account the ASER sampling design. We present two interpretable measures of precision.

- Margin of error – i.e., the % interval around the point estimate that almost certainly contains the population estimate (i.e., with 95% probability). For instance, if *x* is the margin of error then the population proportion lies within $\pm$ *x*% of the sample proportion with 95% probability. As an illustration, consider Table 1 on learning achievements of children in Classes 3 to 5.

In Table 1, the estimates of the average proportion are in column two. Thus, in Himachal Pradesh, 85% of children in standards 3 to 5 could read a level 1 text or more. The standard error of this estimate is 0.01. The 95% confidence interval is the interval $\pm$0.2 around 0.85. To compare across states with different means, we express the interval as a percentage of the estimate itself. This is the margin of error. In the case of Himachal Pradesh, the margin of error is 2.25% which means that with 95% probability the population proportion of children in class 3-5 who can at least read a Level 1 text lies within $\pm$2.25% of 0.85, i.e., between 0.83 and 0.87.

- 95% confidence bands – we use these in the district-level analysis to see whether the estimates are useful in targeting districts. If variability is high, then estimates may not be useful for targeting.

Suppose $\hat{p}$ is the estimated sample proportion and $\hat{\sigma}$ is the associated standard error. From statistical theory, it is known that the interval [$\hat{p} \pm 2\hat{\sigma}$] contains the population proportion with 95% probability. The *margin of error* expresses the confidence interval in terms of the sample estimate. It is thus defined as

$$me = \frac{2\hat{\sigma}}{\hat{p}} \tag{1}$$

A margin of error of 10% is regarded as an acceptable degree of precision in many studies (United Nations, 2005). Equation (1) says this would be true if $\hat{\sigma} \leq 0.05\hat{p}$, i.e., if the standard error is less than or equal to 5% of the sample proportion. Estimates with a margin of error in excess of 20% are regarded as estimates with low precision.

Note that the margin of error depends on the standard error and the estimated proportion and the standard of error itself depends on the estimated proportion. For a given sample size, therefore, a lower precision will be associated with a variable which has a lower incidence in the population.

6. Precision of State Level Estimates

Figure 1 presents the margin of error in state level estimates of learning. All the estimates in this report are based on the ASER survey of 2007. STD12LANG is the % children in class 1-2 who can read letters, words or more in own language. STD12MATH is the % children in class 1-2 who can recognize numbers (1-9) or more. STD35LANG is the % of children in class 3-5 who can read level 1 (Std 1) text or more in own language. STD35MATH is the % children in class 3-5 who can subtract or do more.

From the figure it can be seen that learning outcomes are precisely estimated at the state level. The sample size is clearly adequate. The horizontal lines in the figure are drawn at the 3% and 5% level. Note however that learning outcomes in class 3-5 are relatively less precisely estimated. Although the margin of error here is also less than 5%, they are typically higher than the margin of error of learning outcomes for children in class 1-2. For the latter population, the margin of error is less than 3%. We will return to this issue later.

Figure 2 displays the margin of error for the estimates of the following schooling status variables:

- In School: % children in age group 6-14 years who are enrolled in school
- Pre-School: % children in age group 3-4 years who are in Anganwadi or other pre-school.

- Govt School: % children in age group 6-14 years who attend government schools.

It can be seen that schooling status outcomes are precisely estimated at the state level. However, pre-school status for 3-4 year olds is relatively less precisely estimated.

7. Precision of District Level Estimates

Just as we did for state-level averages, we also computed the precision of estimates for each of the variables and for each of the districts in the 7 states that we studied. Table 2 displays the average margin of error across districts within a particular state. Tables 3 and 4 indicate the range of this variable within the state by presenting the minimum and maximum values of the margin of error for each of the variables of interest.

Margins of error are expectedly higher at the district level as compared to the state level. Furthermore, there is considerable variation in estimates within states (across districts). The margin of error for enrollment in private schools is large. However, this is not a cause for worry because incidence of this variable is low, so that a large margin of error will still give confidence interval bands that are relatively tight. Indeed, the proportion enrolled in government schools would have a low margin of error. At the district level, learning outcomes of class 1-2 are relatively more precise as compared to class 3-5 learning levels. This is to be expected because, as explained earlier, a larger proportion of children in class 1-2 are able to achieve the expected learning outcome. Furthermore within the learning outcomes for class 3-5, language outcomes are usually better estimated than math learning levels. Of the 7 states, the margins of error are on the higher side in Assam, Rajasthan and Karnataka

8. Assessing the precision of district estimates

Given the results in Tables 2 – 4, it is clear that if we desire greater precision in district-level estimates, then the sampling design in districts would have to change. Higher precision could be obtained by increasing the number of villages that are sampled and/or stratifying the household sample at the village level. Both of these options would increase sampling cost. It is therefore important to ask how would greater precision of district estimates help? In other words, how can we assess whether the district estimates are precise enough?

The answer to this depends on how the district estimates are to be used. A reasonable supposition is that we would want to use the learning estimates to target districts for intervention. If the survey can identify the "laggard" districts, then resources could be allocated to give greater importance to those districts.

Clearly, the laggard districts would have to be chosen on the basis of a certain norm, i.e., target all districts that have learning outcomes below the chosen norm. Note that if the norm is 100% learning outcomes then precision is not important at all, because as long as even one child in the sample is not reading at the prescribed level the district will need to be targeted. A more general point is that higher is the learning norm, less important is the precision of district estimates.

On the hand, if the norm is some intermediate level of learning, then it becomes important to be able say whether the learning level in a district is significantly different from the norm and, therefore, precision becomes important.

To illustrate such targeting, suppose we want to distinguish between districts that are above the state mean from the districts that are below it. The confidence interval gives the bounds which contain the true population parameter with 95% probability. These can be used to derive the conditions under which a district will be targeted.

Let *s* be a learning norm, *d* is the district level estimate and σ is the standard error of the district estimate. Then it can be shown that the district estimate is significantly below the learning norm with 95% probability if [7]

$$s > d + 1.65\ \sigma \tag{2}$$

Similarly, a district estimate is significantly above the learning norm with 95% probability if

$$s < d - 1.65\ \sigma \tag{3}$$

Suppose the norm is the average proportion in the state and that we want to distinguish districts that are above this norm from those below this norm. In what follows, we show how such a distinction can be made statistically using this norm. As will become clear, the procedure easily accommodates any other norm as well.

Figures 3 – 9 give learning outcomes, along with their confidence bands, at the district level for each of the 7 states under consideration. For each state, there are four panels. The figures in the top 2 panels report the learning outcomes (language and arithmetic) for children in classes 1 and 2 (STD12LANG and STD12MATH) while the figures in the bottom panels report the learning outcomes for children in standards 3 to 5 (STD35LANG and STD35MATH). The district averages for the learning outcomes can be read off the vertical axis. The horizontal axis plots the districts which are sorted (from lowest to highest) according to the averages in the district. Thus, in each panel, the line in blue is a "learning curve" that plots the learning estimates across districts from low to high. The lines in green and red are the (one-sided) upper and lower confidence limits at the 95% probability level. The horizontal line in black is the state average.

All districts for whom the green line is below the learning norm (the horizontal line in black) are those which are significantly (in a statistical sense) below the state average. All districts for which the red line is above the learning norm are those which are significantly above the state average. The intermediate districts do not satisfy either of these conditions and therefore they cannot be significantly distinguished from the state average.

[7] Notice that we are using one-sided hypothesis tests.

Targeting is good if the intermediate range that is inconclusive is small. Greater precision will reduce the “inconclusive” range and enhance targeting. However, precision alone is not the determinant of good targeting. What also matters is the gradient of the learning curve. Where the curve is steeply sloped, the inconclusive range is small. Where the curve is flat, the inconclusive range will be large despite greater precision of estimates.

To illustrate these points, consider the math learning curve for children in standard 1 and 2 in West Bengal. The learning curve rises sharply upto 0.86 after which the curve is flatter. Districts below 0.78 are significantly below the average and those above 0.93 are above the norm. The inconclusive region comprises those districts with proportions in the range 0.78-0.93. So even though the inconclusive region is large in terms of number of districts, their averages are clustered in the range 0.78-0.93 because the learning curve is flat. The question is whether we want to make fine differences and distinguish between districts in this range. If yes, then greater precision is required. If no, greater precision is not required.

Consider also for the same state the math learning curve for children in standards 3 to 5. The learning curve is steep throughout because sample averages vary greatly across districts. From the figure, it is clear that the inconclusive region is smaller in terms of number of districts. In other words, the data is decisive for targeting. More precision is not necessary.

The figures for Himachal Pradesh reinforce the importance of considering the gradient of the learning curve. From the Tables 2 – 4, it can be seen that the proportion of children in class 1 – 2 that can recognize numbers is more precisely estimated than the proportion of children in class 3 – 5 that can do subtraction. However, the panels for HP show that from the point of view of targeting, it is easier to statistically identify “laggard” districts with respect to math learning levels of the standard 3-5 population than for the standard 1-2 population because the learning curve for the former is steep while that of the latter is flat. Similarly, Karnataka where precision levels are relatively low does well with respect to targeting. In all cases the inconclusive region is small. On the other hand, AP which does relatively well with respect to precision levels does poorly in targeting because of relatively flat learning curves.

To summarize the discussion in the previous sections, state level averages are estimated precisely (within 5% or less of the average with 95% probability). However, even here, learning outcomes in class 3-5 are relatively less precisely estimated. The reason is that the average proportions here are closer to 50% than the learning proportions of children in class 1-2. For a given sample size, the closer the proportion is to 50%, the greater is the variability. Increasing precision would require a larger sample size for children in class 3 – 5. Gains in precision would also be more if increased sample size is allocated to districts with greater variability.

Notice also that the reported learning outcomes of children in class 3-5 involve a higher level test. Therefore, if we report the learning outcomes of a higher level test, variability will increase if the average proportion moves closer to 50%. However, if the test is so stringent that the proportion drops close to zero, then variability will actually drop. But such a test may not be informative and therefore not of interest.

District-level estimates are less precisely estimated compared to state averages. At the 95% probability level, the population proportion lies within 8-18% of the estimate on average. And the precision varies across districts and according to the learning outcome. Once again, learning outcomes of class 1-2 are relatively more precise as compared to class 3-5 learning levels.

Should sample design change to make district estimates more precise? However, greater precision may not be worth the extra cost if the district learning curve is steep. Where the district learning curve is flat targeting would only be possible with high precision. However, is targeting necessary when there is not that much variation between districts? Greater precision may not be valuable in this case. In fact one can go further and say that if the constituent units (districts) within a state do not show much variation, then there is not much point in obtaining disaggregated estimates. The desired level of precision cannot therefore be independent of how the estimates are to be used.

9. Change in Sample Design to Improve Precision

If greater precision of district estimates is desired, then how should the sample design be modified? In principle, the sample size can be increased by increasing the number of villages or by increasing the number of households per village or by a combination of both.

When learning outcomes are correlated within a village, increasing the number of households per village may not lead to a substantial increase in precision. On the other hand, it is well known that sampling cost is much more sensitive to the number of villages surveyed than to the number of households surveyed within a village. There is thus a trade-off between precision and cost.

To know how to strike an optimal balance between precision and cost, it is important to figure out the likely gains in precision as well as the increase in costs from increasing the sample size of villages relative to a sampling strategy that keeps the number of villages fixed but increases the sample size by sampling more individuals within a village.

ASER employs a two-stage clustering design. In the first stage, 30 villages within a district are randomly picked. In the second stage, 20 households within a village are randomly sampled and all the children (aged 5-16) in the selected household are tested.

We consider projections of statistical precision from two experiments. In the first experiment, we increase the number of sampled villages by ten (i.e., from 30 to 40) but keep the number of sampled households per village fixed at 20. Thus, total sample size increases from 600 to 800.

In the second experiment, we keep the number of sampled villages fixed at 30. However, we increase the number of sampled households within the village from 20 to 27. The resulting sample size of 810 is comparable to the sample size of 800 in the first experiment.

As a measure of precision, we continue to focus on the *margin of error*. It is the % interval around the point estimate that contains the population estimate almost certainly (i.e., with 95% probability). For instance, if *x* is the margin of error then the population proportion lies within $\pm x\%$ of the sample proportion with 95% probability. It is defined in equation (1) and we re-iterate it here for continuity.

Suppose $\hat{p}$ is the estimated sample proportion and $\hat{\sigma}$ is the associated standard error. From statistical theory, it is known that the interval [$\hat{p} \pm 2\hat{\sigma}$] contains the population proportion with 95% probability. The *margin of error* expresses the confidence interval in terms of the sample estimate. It is thus defined as

$$me = \frac{2\hat{\sigma}}{\hat{p}}$$

10. Experiment I

In Experiment I we increase the number of villages per district from 30 to 40, keeping the number of sampled households within each village unchanged at 20. Therefore, Experiment I involves increasing the sample size from 600 to 800 households per districts.

Let 0 index the pre-experiment variables and 1 index the variables post Experiment I. Suppose we estimate the proportion of children in the age-class group *X* who satisfy the characteristic *y*. Denote the population proportion by *p*. The variance of the estimate (pre-experiment) is given by:

$$V_0 = \frac{d_0\, p(1-p)}{N_0 - 1}$$

where *d* is the design effect and N_0 is the total number of sampled children in the age-class group *X*. The design effect $d = 1 + (\bar{b} - 1)\rho$ where *b* is the average number of sampled children per village and ρ is the intra-cluster correlation effect.

The estimates of standard error and the margin of error are

$$\hat{\sigma}_0 = \sqrt{\frac{d_0\, \hat{p}(1-\hat{p})}{N_0 - 1}}$$

$$me_0 = \frac{2\hat{\sigma}_0}{\hat{p}} \tag{4}$$

Now imagine that we change the number of sampled villages to 40 but do not change the sample design at the household level. Therefore, the design effect is unaltered since there has been no change in the number of sampled households per village.

Hence, the post-experiment variance is given by

$$V_1 = \frac{d_0\, p(1-p)}{N_1 - 1}$$

where N_1 is the total number of sampled children in the age-class group X under Experiment I. To calculate N_1 we use the data to calculate the average number of children per village in the district. This average is multiplied by ten (the additional villages to be sampled) and added to N_0.

Accordingly

$$\hat{\sigma}_1 = \sqrt{\frac{d_0\, \hat{p}(1-\hat{p})}{N_1 - 1}} \text{ and}$$

$$me_1 = \frac{2\hat{\sigma}_1}{\hat{p}} \tag{5}$$

11. Experiment II

In Experiment II, we keep the number of sampled villages per district unchanged at 30, but increase the number of sampled households within each village from 20 to 27. Therefore, Experiment II increases the sample size from 600 to 810 households per district. This is comparable to the increase in Experiment I, though the entire increase comes from increasing households within villages.

As before, the margin of error of the estimates from the survey is given by (4) namely:

$$me_0 = \frac{2\hat{\sigma}_0}{\hat{p}}$$

Let 2 denote the variables in Experiment II. As the village level sample changes, so will the design effect. From sampling theory, we can write the pre-experiment design effect as

$$d_0 = 1 + (\bar{b}_0 - 1)\rho$$

where $\bar{b}$ is the average number of children in the village (in the relevant age-class group) and ρ is the intra-cluster correlation coefficient.

Hence ρ can be recovered as

$$\rho = \frac{(d_0 - 1)}{\bar{b}_0 - 1}$$

As the intra-cluster correlation coefficient is invariant to sampling design, the design effect in experiment II can be projected as

$$d_2 = 1 + (\bar{b}_2 - 1)\rho$$

Where $\bar{b}_2$ is the average number of children in the village under Experiment II. To derive $\bar{b}_2$, we (i) compute the average number of children per household in a district (ii) multiply it by seven (the additional households surveyed) and (iii) and add the product to the average number of children per village of the pre-experiment sample ($\bar{b}_0$). To calculate the number of children sampled in the new design (denoted as N_2), we simply multiply $\bar{b}_2$, the average number of children in a village under Experiment II, by 30 (the total number of villages in the sample).

The standard error and the margin of error is then

$$\hat{\sigma}_2 = \sqrt{\frac{d_2\,\hat{p}(1-\hat{p})}{N_2 - 1}} \text{ and}$$

$$me_2 = \frac{2\hat{\sigma}_2}{\hat{p}} \qquad (6)$$

12. Findings of Experiments

We continue to report on the precision of the following variables.

- % children in age group 3-4 years who are in Anganwadi or other pre-school.
- % children in age group 6-14 years who are out of school
- % children in age group 6-14 years who are in private school
- % children in class 1-2 who can read letters, words or more in own language
- % children in class 1-2 who can recognize numbers (1-9) or more

- % children in class 3-5 who can read level 1 (Std 1) text or more in own language
- % children in class 3-5 who can subtract or do more

The states chosen for analysis remain Assam, West Bengal, Himachal Pradesh, Bihar, Rajasthan, Andhra Pradesh and Karnataka.

Data from ASER 2008 was used for the analysis. For each district within each state, we computed the margin of error and the projected decline in this variable when the sample design is changed in the manner of Experiment I and Experiment II.

Tables 5-8 summarize the findings. The figures in these tables are state averages of district level estimates i.e., they are a state-wise average measure of the precision of district estimates. STD12-Lang is the % children in class 1-2 who can read letters, words or more in own language. STD12-Math is the % children in class 1-2 who can recognize numbers (1-9) or more. STD35-Lang is the % of children in class 3-5 who can read level 1 (Std 1) text or more in own language. STD35-Math is the % children in class 3-5 who can subtract or do more.

The salient features are the following.

- The margin of error for district-level estimates of the % of children who are out of school are very high, often of the order of 50% or more. This is because the proportion of children out of school is very low (less than 5%) and therefore a precise estimate requires a very large sample. On the other hand, the proportion of children who are enrolled in school is very high and therefore its precision is very high (margin of error < 4%). Therefore the low precision of estimates of children out of school is not of serious concern.

- Among other variables, the % of children enrolled in private schools has the lowest precision. This is probably driven by the fact that the proportion of children in private school is closer to 0.5 than say, those enrolled in any other type of school. The sample size needed to precisely measure parameters in the middle range are far greater than those to estimate parameters whose true values lie in the extremes. In fact, the maximum sample size is required to measure parameters whose true value is 0.5.

- Among the learning outcomes, the outcomes for children in standard 1 and 2 are better estimated than outcomes for children in standards 3 to 5. Note that the learning outcome being considered for Std. 1-2 is more generous than that for Std. 3-5, implying that the population proportion for Std. 1-2 is likely to be much greater than that for Std. 3-5. Therefore, even though the sample sizes for Std. 3-5 learning outcomes are greater, the precision is smaller.

- The district-level precision of learning outcomes for standard 1 and 2 is adequate in most cases. The margin of error is less than 10% in four states and hovers a little above 10% in other cases.

- In almost all cases, Experiment I yields a greater reduction in the margin of error than Experiment II. **Hence ameliorating the adverse effects from cluster design is more important than merely increasing the sample size.**

- However, the reductions in margin of error are not so large that the margin of error is brought below 10% when it was originally above it.

13. Ideal Sample Sizes: The question and methodology

We elaborate on the last point of the earlier section. As reductions in margin of error from the experiments are not very large, it must be that we must consider even more drastic changes to reduce the margin of error at the district level to below 10%. Because Experiment I dominates Experiment II in almost all cases, we consider the sample sizes with respect to Experiment 1 that are likely to yield estimates with acceptable precision. Since Experiment I does not result in changes in the design effect, all the subsequent calculations are done using pre-experiment values of the estimates.

Two scenarios are considered. In the first scenario, we ask what the sample size should be so that the estimates have a margin of error of 5%. In the second case, we compute the sample size that would yield estimates with a margin of error of 10%.

The methodology is straightforward. From the definition of margin of error, we can write

$$me = \frac{2\sigma}{p} = \frac{2\sqrt{\dfrac{d\,p(1-p)}{N-1}}}{p}$$

When the margin of error is restricted to be 5%, we must have

$$\frac{2\sqrt{\dfrac{d\,p(1-p)}{N-1}}}{p} = 0.05 \text{ or}$$

$$\frac{4d\,p(1-p)}{N-1} = 0.0025\,p^2$$

Hence the sample size that satisfies the above would be

$$N = 1 + \frac{4d\,p(1-p)}{0.0025\,p^2}$$

Similarly the sample size that leads to estimates with 10% margin of error is

$$N = 1 + \frac{4d\, p(1-p)}{0.01\, p^2}$$

It can be seen that the sample size that yields estimates with 5% margin of error would be four times larger than sample size with 10% margin of error.

The results are summarized in Tables 9-12. As before, we report on the state averages of the district-level sample size. The first column reports the sample size in ASER 2008. The second column reports the state average of the sample size needed at the district level to generate estimates with 5% margin of error. The third column reports the corresponding figure for a 10% margin of error.

Once again the figures for the proportion of children out of school can be disregarded. The proportion of children in school is precisely estimated by ASER; indeed the tables show that even a smaller sample will generate estimates with 5% and 10% margin of error.

For other variables, a larger sample will be necessary to obtain estimates with 5% margin of error. As noted earlier, samples that are accurate to the extent of 5% of margin error have to be four times larger than samples that allow 10% margin of error. Therefore, in instances, where the margin of error is greater than 10%, the required increase in sample size (and hence cost) is more than four times the existing sample. Thus achieving such a level of precision will be very costly.

On the other hand, a 10% precision level is achievable in many cases with a relatively modest increase in sample size. Learning outcomes at the standard 1-2 level is already estimated with close to 10% precision level in Andhra Pradesh, Himachal Pradesh, Karnataka and Rajasthan and so no changes in sample size are necessary here. Assam, Bihar and Rajasthan need about a doubling of sample size to reach the 10% precision level.

It is the estimates of learning outcomes for the standard 3-5 group that uniformly need larger sample sizes to achieve the 10% precision level. Even here, however, the required sample sizes for Andhra Pradesh, Himachal Pradesh, Karnataka and Rajasthan are lower than for Assam, Bihar and Rajasthan.

14. A Bootstrap Study

The results presented in the earlier sections are based on a methodology extrapolates the margin of error to larger sample sizes based on the ASER 2008 sample design and data. The extrapolation uses two key assumptions. First, if the sample size increases (by either route of Experiment I or Experiment II), we need to work out the likely increase in the sample of the relevant age-class group. For instance, consider Experiment I, where we examine the consequences of adding 10 more villages to the ASER sample. To get the new sample size we calculated the average number of children in a village per district from ASER 2008, multiplied it by 10 and added it to the original sample size. If we were to actually add 10 more villages, we might get a sample size which is quite different than what has been extrapolated. Second, the standard of error and hence the margin of error depends on the estimated proportion which is subject to sampling variability.

To check the robustness of our findings, we considered an alternative methodology that is entirely free of these assumptions. However, it is data intensive and therefore is not feasible to implement as widely as the first methodology.

The alternative methodology consists of bootstrapping the empirical distribution of the parameter of interest (e.g., the proportion of children in standards 1-2 who who can read letters, words or more in own language) for different sample sizes. This was made possible by a pilot survey in Vaishali district of Bihar in April 2009. 90 villages were independently sampled using PPS from the Census Village Directory and an ASER survey was conducted in each of these 90 villages by sampling 20 households in each of the selected villages. In other words, we replicated ASER in 90 villages in a single district.

With an independent sample of 90 villages, we can derive the empirical distribution of each of 7 parameters at the district level under alternative sample sizes chosen here to range from 30 to 60. The mean of the empirical distribution is an unbiased estimate of the true population proportion and its standard deviation is an unbiased estimate of the unobserved standard error.

The empirical distribution is derived by using a bootstrap. Bootstrapping generates many estimates of a single statistic that would ordinarily be calculated from one sample. As stated earlier, ASER provides one sample from which one estimate of the proportion and its standard error can be derived. Therefore, it does not tell us how these statistics vary. Using the bootstrap methodology, we randomly extract a new sample of n villages out of the 90 sampled villages. The villages are drawn without replacement. By doing this repeatedly – 5000 bootstrap iterations – we create 5000 samples with n villages and (n X 20) households, and compute the proportion for each of these samples. Thus we get an estimate of the distribution of the statistic. We then use the mean and standard deviation of this distribution as an estimate of the population proportion, p and its standard error, σ, to calculate the margin of error. Finally, we repeat these 5000 iterations for each of the 7 proportions for a sample size of 30, 40, 50 and 60 villages.

To summarize, the bootstrap procedure involves the following steps. For each of the 7 variables of interest:

1. Draw 30 villages without replacement from the 90 sampled villages – this gives us a sample size of 600 (=30 X 20) households for the district which is the ASER sample size.

2. Calculate the proportion of interest.
3. Repeat Steps 1 and 2, 5000 times to get the empirical distribution of *p*.
4. Calculate the mean and standard deviation of the distribution and use these to calculate the margin of error.
5. Repeat Steps 1 – 4 with draws of 40 (sample size = 800 households), 50 (sample size of 1000 households) and 60 (sample size = 1200 households) villages.

Results from the bootstrap are presented in Table 13. For each of the 7 variables of interest we present the mean of the empirical distribution (*p*), its standard error (*se(p)*) and the associated margin of error (*me(p)*). This is done under 4 scenarios – the original sample size of ASER of 30 villages X 20 households, 40 villages X 20 households – the sample size in Experiment I, 50 villages X 20 households and 60 villages X 20 households.

Note that the estimate of *p* in all cases is very robust. This is not surprising since it is an unbiased estimator of the underlying population proportion. Also as expected, the standard error falls as the sample size increases as does the margin of error. The bootstrap experiment seems to imply a simple rule: Double the sample size (by doubling the number of clusters) to halve the margin of error. Note that even this will not lead to a 5% margin of error at the district level in most cases, unless the initial margin of error was around 10% as is the case with Std12-Math. Learning outcomes are fairly well estimated at the district level with increasing the sample size by 10 clusters often giving reasonable margins of error. Private school proportions remain problematic.

Overall, these results point in the same direction as the earlier findings. Reducing the margin of error to 5% level is very expensive. Reducing it to the 10% level could be manageable. Given the tradeoff between precision and cost, it becomes important, therefore, to figure out what the estimates will be used for. If the objective is to identify districts for action, the current precision levels may be sufficient. Further, high margins of error are often associated with low incidence. For instance, the worst offender is the private school proportion. In our example this proportion is 8.2% with a margin of error of 35%. What this implies is that with 95% probability the true proportion lies between 5.3% and 11.1%. This band though wide in relative terms is not that wide in absolute terms and may be sufficient to identify Vaishali as a low private school district.

15. Concluding Remarks

- State level averages are estimated precisely (the 95% confidence band lies within 5% or less of the estimate). Learning outcomes of children in class 3-5 are relatively less precisely estimated.
- District-level estimates are less precisely estimated compared to state averages. At the 95% probability level, the population proportion lies within 8-18% of the estimate on average. The precision varies across districts and according to the learning outcome. Once again, learning outcomes of class 1-2 are relatively more precise as compared to class 3-5 learning levels.

- District-level estimates can be made more precise by increasing the sample size. This can be done either by increasing the number of sampled villages or by increasing the number of households in the sampled villages. The former is a more effective way of increasing precision. It is also more costly.
- How much the sample size needs to be increased depends on the desired level of precision. A 10% precision level is achievable in many cases with a relatively modest increase in sample size especially for learning outcomes for children in the classes 1 – 2.
- The sample size that yields estimates with 5% margin of error is four times larger than sample size with 10% margin of error. Achieving such precision is therefore very costly.
- Should sample design change to make district estimates more precise? It depends very much on how the estimates are to be used. For instance, if these estimates are to be used for targeting resources at lagging districts, then the existing precision is in fact sufficient to discriminate between 'high' outcome and 'low'outcome districts. On the other hand, where districts are tightly bunched together, then the existing precision fails to discriminate with enough statistical power. However, such a scenario may not call for targeting in any case.

References

United Nations (2005), *Designing Household Survey Samples: Practical Guidelines*, Studies in Methods, Series F No. 98, Department of Economic and Social Affairs, Statistics Division.

Table 1: State Level Learning in Classes 3 – 5

State	% of Children in Std. 3-5 who can read Level 1 text or more		
	Point Estimate	**Standard Error**	**Margin of error (%)**
Assam	0.66	0.013	3.76
West Bengal	0.77	0.010	2.66
Himachal Pradesh	0.85	0.010	2.25
Bihar	0.69	0.009	2.46
Rajasthan	0.58	0.009	3.08
Andhra Pradesh	0.75	0.009	2.27
Karnataka	0.57	0.009	3.20
All India	0.66	0.002	0.69

Table 2: District level Estimates of Margin of Error: State Average

	% in Pre-School	% in School	% in private school	STD12 LANG	STD12 MATH	STD35 LANG	STD35 MATH
Assam	27.82	5.89	50.67	9.94	10.47	17.42	19.92
W. Bengal	11.07	2.41	74.40	8.63	7.10	12.85	13.14
Himachal	7.57	0.95	43.95	6.30	5.56	7.55	9.54
Bihar	27.32	3.62	55.94	11.68	12.80	13.14	13.22
Rajasthan	32.84	3.18	36.41	14.40	14.20	16.95	20.55
Andhra	17.69	2.34	30.16	9.89	8.25	11.04	14.26
Karnataka	5.82	1.69	49.31	7.82	8.38	15.20	21.36

Table 3: District Level Estimates of Margin of Error: District Minimum

	% in Pre-School	% in School	% in private school	STD12 LANG	STD12 MATH	STD35 LANG	STD35 MATH
Assam	4.62	1.26	29.00	3.95	2.98	8.89	7.74
W. Bengal	0.00	0.92	27.20	0.00	1.37	1.29	1.34
Himachal	2.67	0.00	19.58	1.01	1.37	2.77	4.12
Bihar	0.93	0.62	31.63	6.22	4.37	6.57	4.40
Rajasthan	15.13	0.91	16.09	6.79	5.89	10.45	7.63
Andhra	6.41	1.01	15.71	4.29	3.06	3.89	5.93
Karnataka	0.00	0.24	23.90	1.25	1.89	5.06	5.63

Table 4: District Level Estimates of Margin of Error: District Maximum

	% in Pre-School	% in School	% in private school	STD12 LANG	STD12 MATH	STD35 LANG	STD35 MATH
	Maximum of Margin of Error (95%)						
Assam	76.81	42.56	111.65	23.60	19.04	29.15	41.55
W. Bengal	20.62	7.28	131.18	31.86	14.23	24.96	24.93
Himachal	18.31	3.81	75.17	13.21	12.21	12.60	16.15
Bihar	86.68	16.63	132.87	20.61	21.97	25.81	25.08
Rajasthan	52.16	6.35	76.88	22.19	25.02	26.51	33.00
Andhra	35.79	7.35	42.64	17.35	14.56	20.02	23.28
Karnataka	19.34	6.18	150.82	15.27	16.43	24.71	36.14

Table 5: State-averages of precision of district estimates – Andhra Pradesh & Assam

	Andhra Pradesh			Assam		
	2008 Survey	Experiment I	Experiment II	2008 Survey	Experiment I	Experiment II
% in Pre-School	11.15	9.28	9.20	22.35	18.48	19.01
% in School	2.00	1.73	1.88	3.04	2.63	2.88
% out of School	58.09	50.14	53.90	53.37	45.99	49.97
% in Private School	29.64	25.59	28.81	42.69	36.79	41.05
STD12-Lang	7.43	6.40	6.84	13.54	11.61	12.74
STD12-Math	7.11	6.12	6.51	12.65	10.84	11.86
STD35-Lang	10.56	9.11	9.78	18.80	16.16	17.84
STD35-Math	13.59	11.72	12.68	27.09	23.26	25.69

Table 6: State-averages of precision of district estimates –Bihar & Himachal Pradesh

	Bihar			Himachal Pradesh		
	2008 Survey	Experiment I	Experiment II	2008 Survey	Experiment I	Experiment II
% in Pre-School	22.18	18.72	19.80	13.08	10.86	10.82
% in School	2.73	2.34	2.54	2.16	1.86	2.03
% out of School	55.28	47.31	50.95	59.24	51.11	55.42
% in Private School	52.41	44.84	49.04	30.68	26.46	29.76
STD12-Lang	12.28	10.47	11.29	8.55	7.37	7.92
STD12-Math	12.02	10.24	11.04	8.02	6.91	7.38
STD35-Lang	12.01	10.29	11.18	11.98	10.33	11.21
STD35-Math	14.71	12.59	13.73	15.44	13.31	14.53

Table 7: State-averages of precision of district estimates –Karnataka & Rajasthan

	Karnataka			Rajasthan		
	2008 Survey	Experiment I	Experiment II	2008 Survey	Experiment I	Experiment II
% in Pre-School	8.56	7.36	7.97	31.90	25.86	25.63
% in School	1.68	1.44	1.53	3.79	3.27	3.63
% out of School	74.80	64.39	65.40	50.18	43.25	47.55
% in Private School	36.36	31.26	34.54	33.18	28.59	32.21
STD12-Lang	7.78	6.69	6.94	14.69	12.60	13.64
STD12-Math	8.56	7.37	7.81	14.56	12.47	13.42
STD35-Lang	13.97	12.03	12.89	15.64	13.44	14.76
STD35-Math	22.86	19.67	21.16	21.78	18.71	20.52

Table 8: State-averages of precision of district estimates – West Bengal

	West Bengal		
	2008 Survey	Experiment I	Experiment II
% in Pre-School	16.36	13.84	14.33
% in School	3.07	2.65	2.88
% out of School	55.12	47.50	51.04
% in Private School	60.88	52.53	56.63
STD12-Lang	9.83	8.45	9.14
STD12-Math	9.41	8.09	8.75
STD35-Lang	13.44	11.54	12.35
STD35-Math	17.81	15.30	16.42

Table 9: ASER sample sizes and Ideal Sample Sizes: Andhra Pradesh and Assam

	Andhra Pradesh			Assam		
	ASER 2008	For 5% ME	For 10% ME	ASER 2008	For 5% ME	For 10% ME
% in Pre-School	67	411	103	112	2665	666
% in School	841	158	39	830	404	101
% out of School	841	117667	29417	830	100951	25238
% in Private School	841	31727	7932	830	67523	16881
STD12-Lang	188	491	123	239	1798	450
STD12-Math	188	424	106	239	1566	392
STD35-Lang	286	1417	354	282	4044	1011
STD35-Math	286	2312	578	281	8478	2119

Table 10: ASER sample sizes and Ideal Sample Sizes: Bihar and Himachal Pradesh

	Bihar			Himachal Pradesh		
	ASER 2008	For 5% ME	For 10% ME	ASER 2008	For 5% ME	For 10% ME
% in Pre-School	177	4283	1071	63	501	125
% in School	1143	406	101	875	191	48
% out of School	1143	157033	39258	875	128115	32029
% in Private School	1143	133000	33250	875	35242	8810
STD12-Lang	398	2576	644	192	642	160
STD12-Math	397	2425	606	191	526	131
STD35-Lang	431	2642	661	300	1859	465
STD35-Math	429	3891	973	300	2963	741

Table 11: ASER sample sizes and Ideal Sample Sizes: Karnataka and Rajasthan

	Karnataka			Rajasthan		
	ASER 2008	For 5% ME	For 10% ME	ASER 2008	For 5% ME	For 10% ME
% in Pre-School	117	428	107	97	4338	1084
% in School	753	203	51	1017	742	185
% out of School	753	198522	49631	1017	108199	27050
% in Private School	753	43630	10907	1017	52906	13227
STD12-Lang	171	516	129	267	2439	610
STD12-Math	171	577	144	264	2388	597
STD35-Lang	270	2285	571	370	3868	967
STD35-Math	269	6062	1515	368	7586	1897

Table 12: ASER sample sizes and Ideal Sample Sizes: West Bengal

	West Bengal		
	ASER 2008	For 5% ME	For 10% ME
% in Pre-School	100	1173	293
% in School	686	356	89
% out of School	686	92368	23092
% in Private School	686	105450	26363
STD12-Lang	190	919	230
STD12-Math	190	838	210
STD35-Lang	252	1938	484
STD35-Math	252	3387	847

Table 13: Margin of Error under Different Sample Sizes
Estimates from Vaishali District – 5000 Bootstrap Iterations

	% in Pre-School			% Out of School		
	p	*se(p)*	*me(p)*	*p*	*se(p)*	*me(p)*
30 villages	0.378	0.040	21.37	0.119	0.015	25.14
40 villages	0.377	0.033	17.36	0.120	0.012	19.46
50 villages	0.377	0.026	13.77	0.120	0.009	15.35
60 villages	0.378	0.021	10.91	0.120	0.007	12.40

	% in Private School		
	p	*se(p)*	*me(p)*
30 villages	0.082	0.015	35.80
40 villages	0.082	0.011	27.71
50 villages	0.082	0.009	22.28
60 villages	0.082	0.007	17.84

	STD12-Lang			STD12-Math		
	p	*se(p)*	*me(p)*	*p*	*se(p)*	*me(p)*
30 villages	0.364	0.029	16.08	0.536	0.028	10.57
40 villages	0.363	0.023	12.76	0.536	0.023	8.39
50 villages	0.363	0.018	9.99	0.536	0.018	6.62
60 villages	0.363	0.015	8.04	0.536	0.014	5.25

	STD35-Lang			STD35-Math		
	p	*se(p)*	*me(p)*	*p*	*se(p)*	*me(p)*
30 villages	0.468	0.027	11.58	0.389	0.029	14.97
40 villages	0.468	0.021	9.10	0.388	0.023	12.07
50 villages	0.468	0.017	7.33	0.388	0.019	9.64
60 villages	0.468	0.013	5.77	0.388	0.015	7.65

Figure 1: Precision of Learning Outcomes at State-level

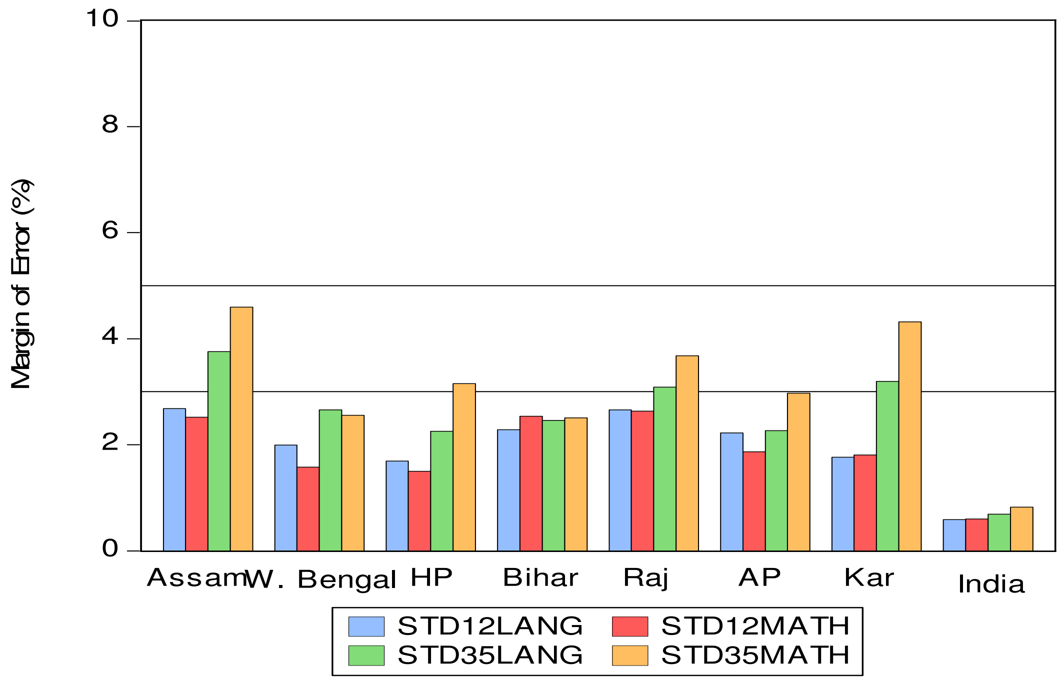


Figure 2: Precision of State-level Estimates of Schooling Status

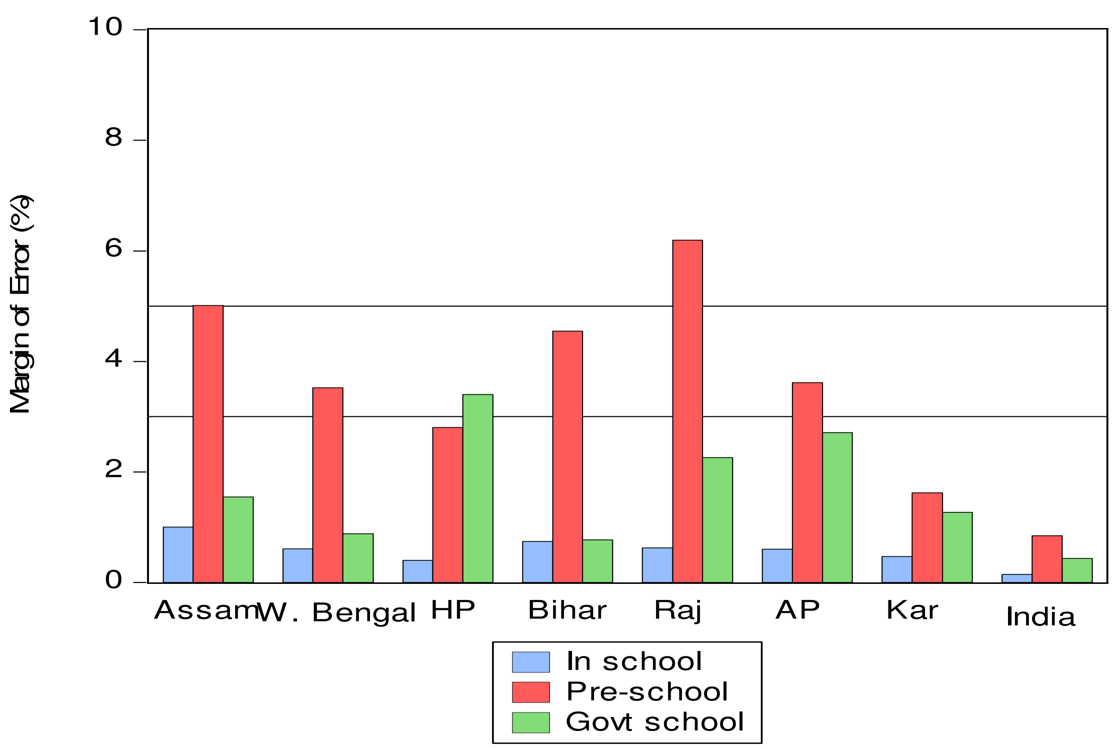


Figure 3: Precision of District Learning Estimates: Assam

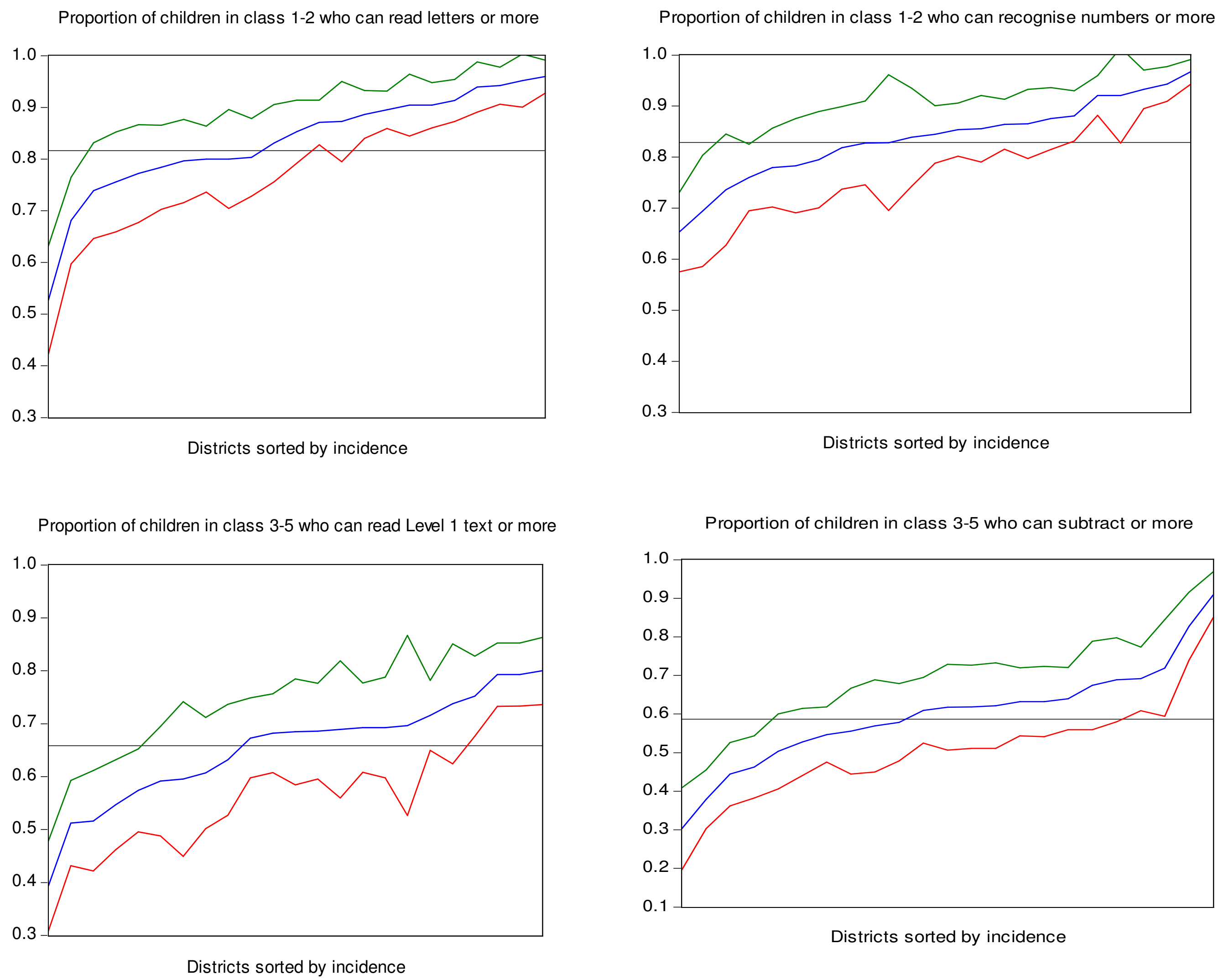
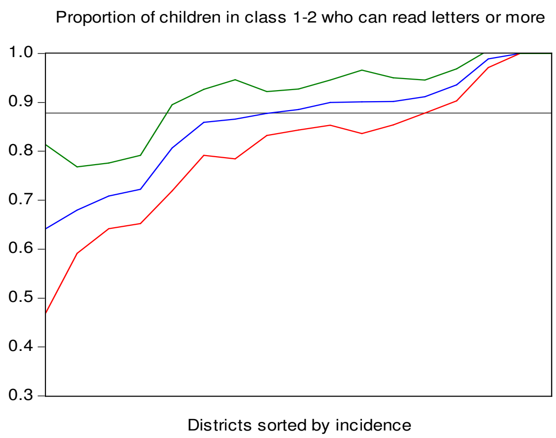
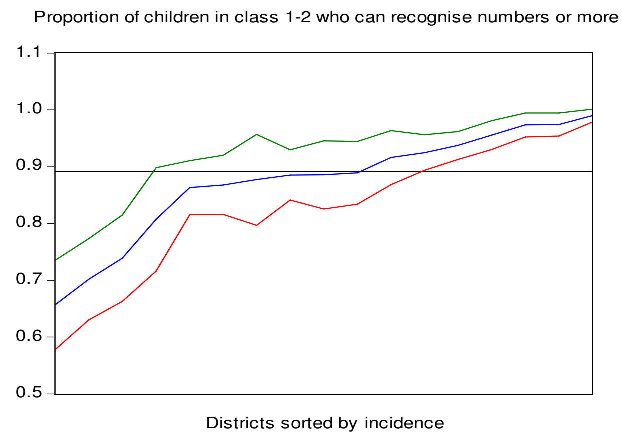
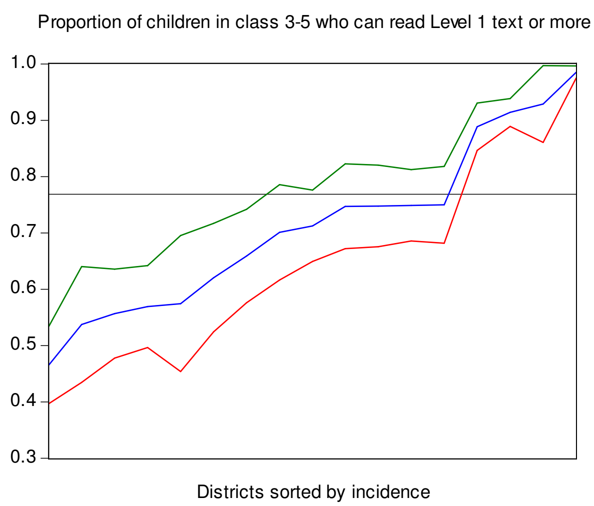


Figure 4: Precision of District Learning Estimates: West Bengal







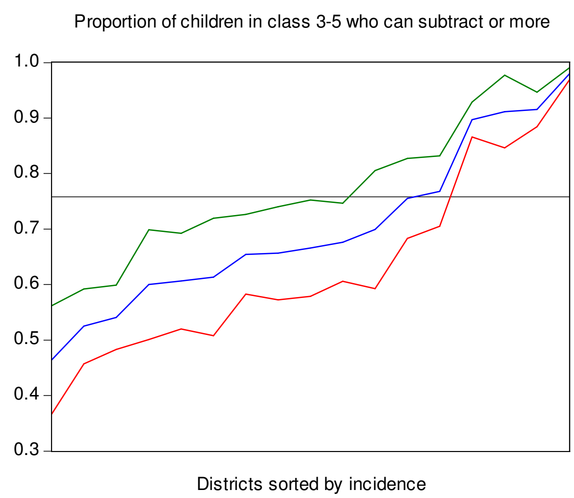


Figure 5: Precision of District Learning Estimates: Himachal Pradesh

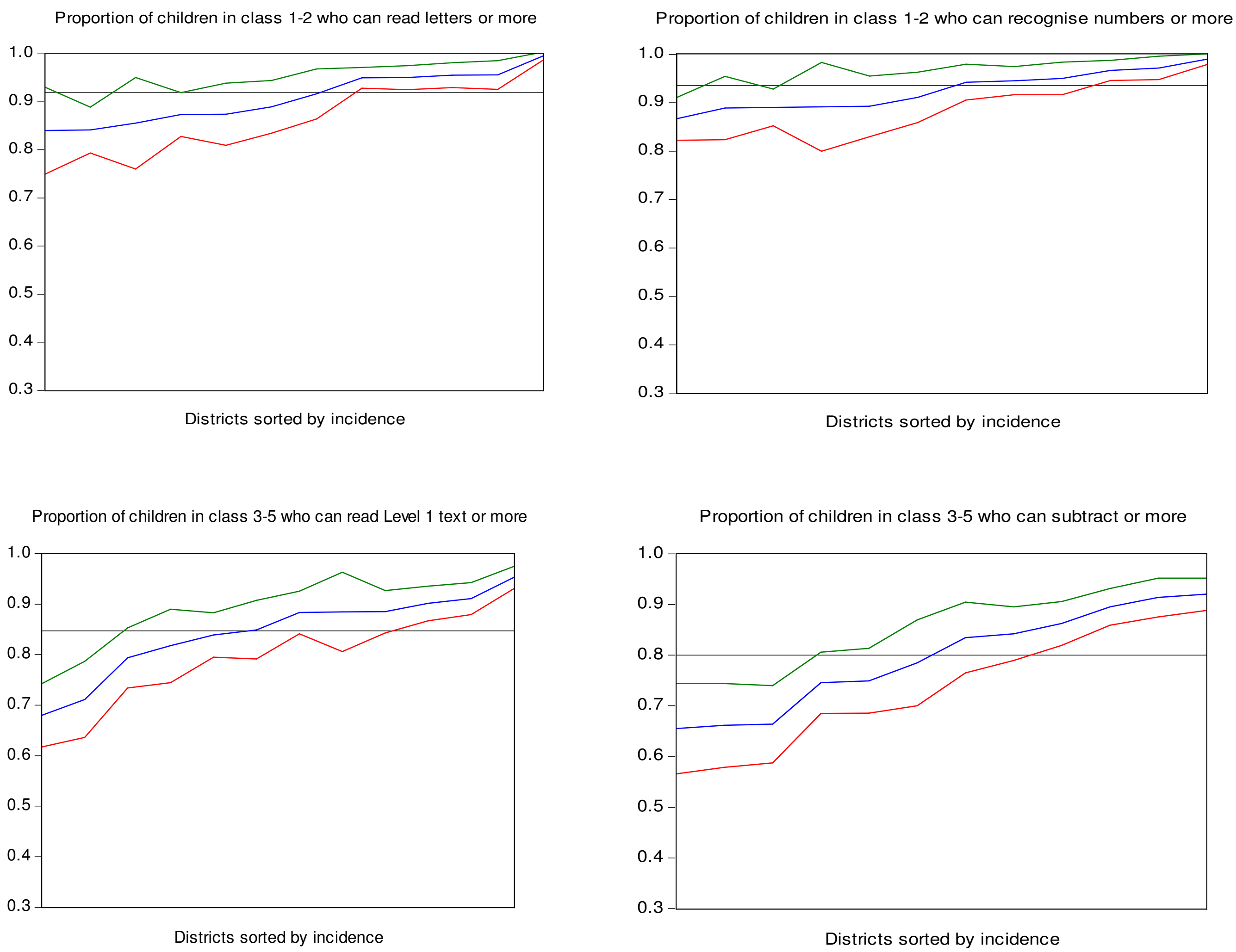


Figure 6: Precision of District Learning Estimates: Bihar

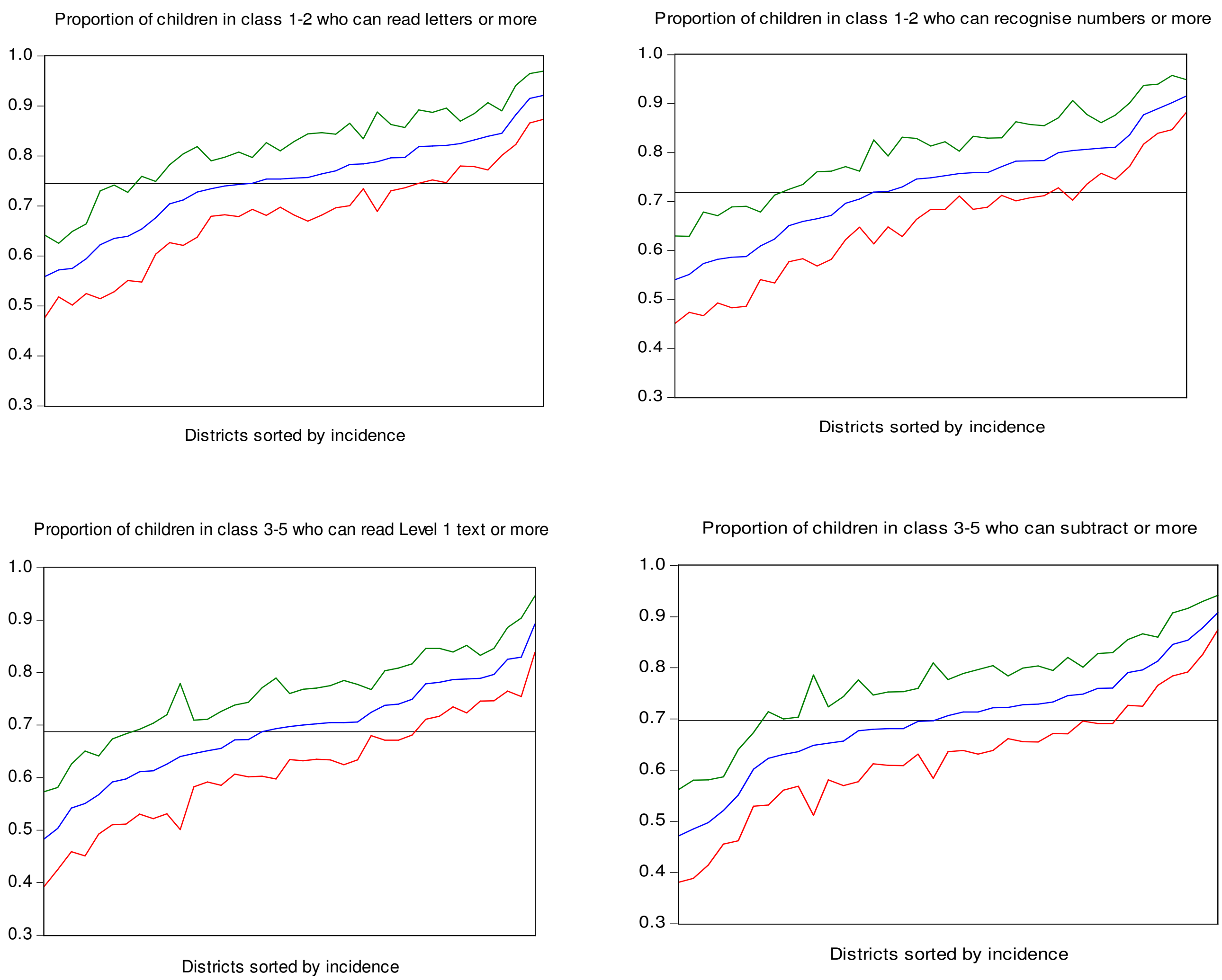


Figure 7: Precision of District Learning Estimates: Rajasthan

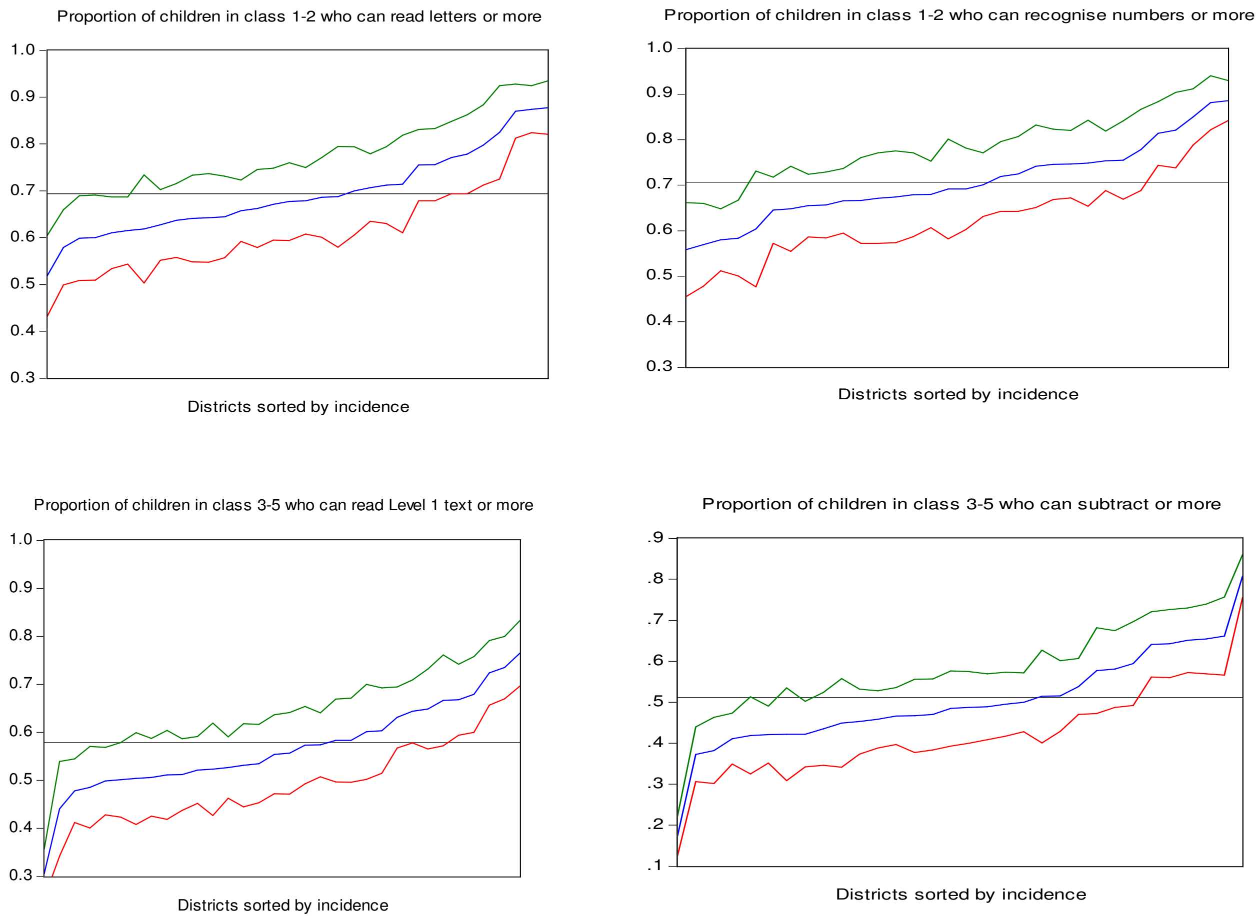


Figure 8: Precision of District Learning Estimates: Andhra Pradesh

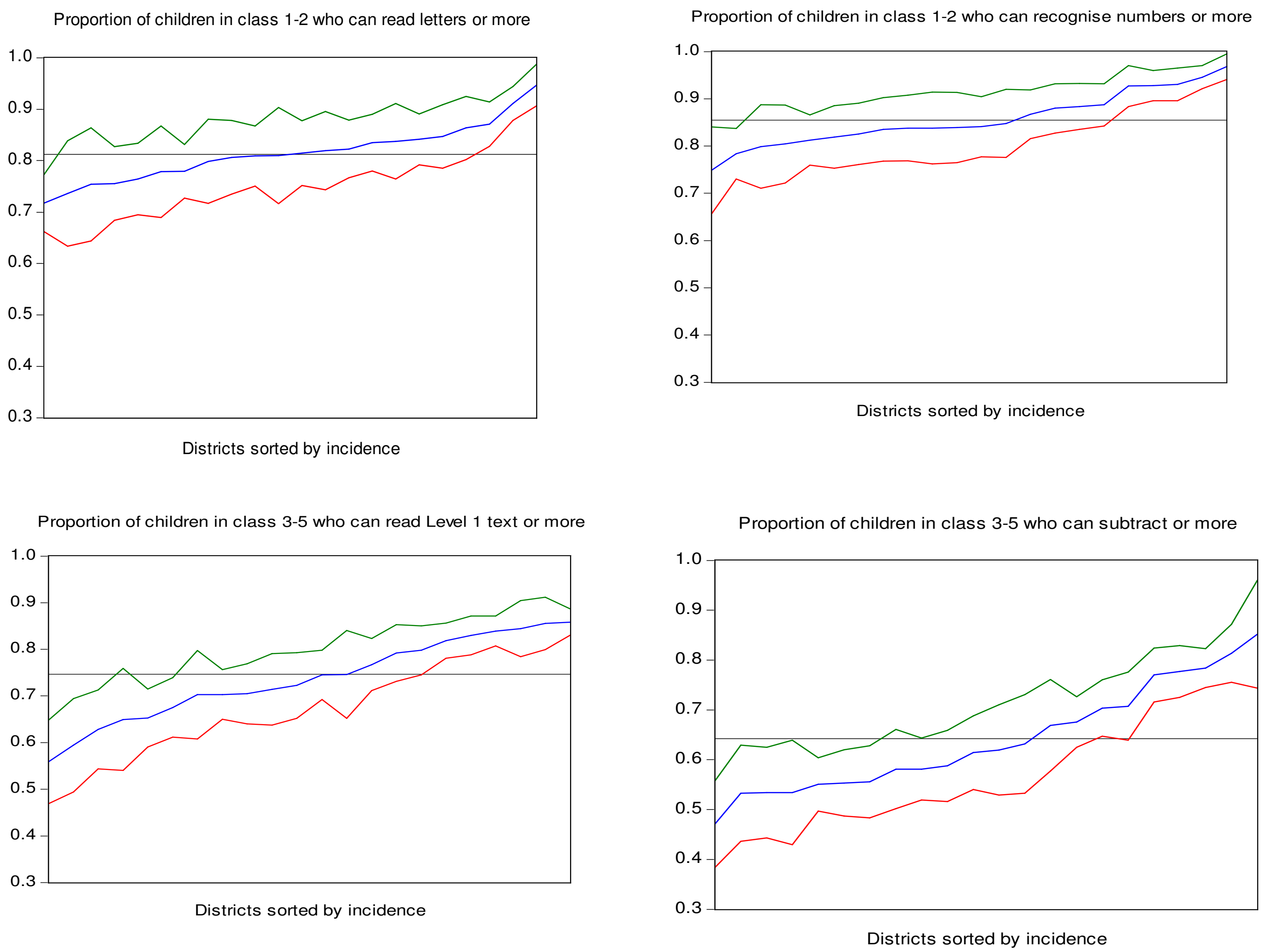


Figure 9: Precision of District Learning Estimates: Karnataka