Annexures

Comparison of age- class composition in sample frames of 2006 and 2007

Age-wise composition of classes

2006

Class studying in	Age of children												
	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	21.9	47.6	17.1	7.9	1.9	2.0	1.7						100
2	2.8	15.1	37.6	28.0	6.2	6.0	1.3	1.7	1.3				100
3	0.7	3.2	10.7	42.3	21.8	12.4	2.7	3.6	1.3	1.5			100
4	1.2		2.8	13.5	31.3	33.0	6.7	6.9	2.1	1.4	1.1		100
5	1.4			3.9	7.0	42.5	21.6	13.5	4.5	2.9	1.6	1.1	100
6	1.9				2.0	11.4	29.0	36.4	10.2	5.5	2.4	1.3	100
7	1.6					3.5	6.9	40.4	27.5	12.4	5.4	2.3	100
8	2.3						1.9	12.6	35.1	30.6	11.3	6.2	100
9	2.0							3.3	11.1	42.5	30.6	10.5	100
10	2.1								2.0	12.3	44.3	39.3	100
11	5.6									3.8	20.9	69.7	100
12	13.3										13.0	73.7	100
Total	**3.5**	**8.9**	**8.6**	**11.6**	**8.0**	**12.5**	**7.3**	**11.2**	**7.9**	**8.0**	**6.9**	**5.5**	**100**

2007

Class studying in	Age of children												
	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	27.1	45.3	16.9	6.5	1.6	1.3	1.3						100
2	4.1	14.4	41.7	27.4	5.4	4.1	1.0	2.0					100
3	1.1	3.4	11.2	44.5	23.6	10.1	2.3	2.5	1.5				100
4	1.6		2.9	13.1	35.2	32.4	6.4	5.3	1.5	1.6			100
5	1.4			3.7	7.4	44.3	23.9	12.0	3.7	2.0	1.6		100
6	1.8				2.2	11.4	31.6	37.2	9.4	4.3	1.5	0.7	100
7	1.6					3.4	7.2	41.3	29.4	11.5	4.1	1.5	100
8	2.4						2.0	12.3	36.5	31.3	10.7	4.7	100
9	2.3							3.3	11.0	45.2	29.6	8.6	100
10	2.3								1.8	12.3	46.1	37.4	100
11	4.2									4.3	21.9	69.7	100
12	16.2										11.9	71.9	100
Total	**4.4**	**8.6**	**9.5**	**12.3**	**9.4**	**13.0**	**8.1**	**11.0**	**7.6**	**7.1**	**5.3**	**3.7**	**100**

Proportion of children of same age in different classes

2006

Class studying in	Age of children												
	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	85.1	72.8	27.0	9.3	3.3	2.1	0.8	2.4	3.0	3.8	5.3	7.2	13.6
2	9.8	20.4	52.7	29.3	9.4	5.8	2.1						12.1
3	5.1	4.2	14.6	43.2	32.0	11.7	4.3	3.8					11.8
4		2.6	3.5	12.5	41.5	28.1	9.7	6.5	2.9				10.7
5			2.3	3.9	10.1	39.6	34.3	14.1	6.6	4.3			11.7
6				1.9	2.4	8.8	37.9	31.1	12.3	6.5	3.4		9.6
7					1.4	2.4	8.0	30.8	29.6	13.1	6.7	3.6	8.5
8						1.6	2.1	9.1	35.7	30.6	13.3	9.1	8.1
9							0.8	1.8	8.4	31.6	26.7	11.3	6.0
10								0.5	1.5	9.3	39.2	43.1	6.1
11									0.2	0.7	4.2	17.5	1.4
12										0.2	1.2	8.2	0.6
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	100

2007

Class studying in	Age of children												
	5	6	7	8	9	10	11	12	13	14	15	16	Total
1	82.2	70.5	23.6	7.0	2.2	1.4	0.5	1.7	2.1	3.0	4.6	6.0	13.3
2	11.8	21.4	55.7	28.3	7.3	4.1	1.5						12.7
3	6.0	5.1	15.1	46.6	32.3	10.0	3.7	2.9					12.9
4		3.0	3.6	12.8	44.9	29.9	9.4	5.8	2.4				12.0
5			2.1	3.7	9.8	42.4	36.5	13.6	6.0	3.5			12.4
6				1.7	2.3	8.6	38.2	33.3	12.1	5.9	2.8		9.8
7					1.1	2.3	7.6	32.4	33.0	14.0	6.6	3.4	8.6
8						1.4	1.9	8.4	36.0	33.2	15.1	9.6	7.5
9							0.7	1.5	7.3	32.1	28.0	11.7	5.1
10								0.4	1.0	7.7	38.2	44.7	4.4
11									0.2	0.6	3.9	17.9	1.0
12										0.1	0.8	6.8	0.4
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	100

The class of the child can be recorded accurately than the age of the child. More peple tend to the age in even numbers

STATE	Districts		Surveyed		Number of children in age group										Number of schools	
	Actual	Surveyed	Villages	House-holds	3-16 years	3-5 years			6-14 years			15-16 years			having classes..	
						Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	1-7/8	1-4/5
Jammu&Kashmir	14	14	379	7669	18627	2651	1387	1264	13800	7595	6205	2176	1263	913	151	109
Himachal Pradesh	12	12	344	6250	13029	2517	1333	1184	9190	4702	4488	1322	657	665	24	207
Punjab	19	19	536	10715	21431	3512	1981	1531	15516	8770	6746	2403	1391	1012	48	358
Uttaranchal	13	13	369	6986	15471	2828	1509	1319	11134	5893	5241	1509	776	733	14	306
Haryana	20	20	560	11433	26010	4877	2802	2075	18187	10240	7947	2946	1696	1250	79	302
Rajasthan	32	32	959	18934	49329	8076	4392	3684	35772	20429	15343	5481	3395	2086	449	365
UP	69	69	2039	40285	107665	17202	9330	7872	80299	45509	34790	10164	5896	4268	92	1,796
Bihar	37	37	1085	21716	58876	11317	6032	5285	43684	24787	18897	3875	2434	1441	454	442
Sikkim	4	1	26	549	916	130	62	68	664	336	328	122	63	59	0	1
Arunachal Pradesh	13	13	326	6184	13991	2690	1388	1302	9703	5170	4533	1598	935	663	96	133
Nagaland	11	10	243	5348	12423	2504	1345	1159	8545	4694	3851	1374	768	606	21	178
Manipur	9	9	236	4911	11153	2204	1198	1006	7806	4105	3701	1143	600	543	34	97
Tripura	4	3	46	955	1811	317	182	135	1299	683	616	195	105	90	22	13
Meghalaya	7	6	149	2835	7444	1458	759	699	5066	2587	2479	920	466	454	9	93
Assam	23	23	635	12426	26323	4550	2302	2248	19335	10125	9210	2438	1360	1078	35	449
West Bengal	17	17	487	9842	18606	3292	1738	1554	13959	7659	6300	1355	753	602	8	354
Jharkhand	22	22	628	12693	31747	6203	3234	2969	23293	12577	10716	2251	1348	903	275	224
Orissa	30	30	875	16300	32604	6269	3204	3065	23212	12313	10899	3123	1598	1525	303	404
Chhattisgarh	16	15	435	8429	19452	3274	1635	1639	13845	7065	6780	2333	1190	1143	77	349
Madhya Pradesh	45	45	1323	26388	61391	10565	5578	4987	45385	24871	20514	5441	3262	2179	343	935
Gujarat	25	25	725	14483	30896	5779	3097	2682	23140	13093	10047	1977	1147	830	503	72
Daman & Diu	2	2	14	1110	2444	471	238	233	1699	897	802	274	131	143	3	8
Dadra & Nagar Haveli	1	1	23	602	1308	219	98	121	939	488	451	150	77	73	17	5
Maharashtra	33	33	971	18448	36378	6759	3637	3122	26107	13808	12299	3512	1874	1638	403	480
Andhra Pradesh	22	22	639	12923	24058	3051	1580	1471	18758	9171	9587	2249	1116	1133	214	365
Karnataka	27	27	790	15625	30283	5165	2654	2511	22731	11425	11306	2387	1198	1189	585	168
Goa	2	2	51	1173	2044	276	140	136	1374	722	652	394	233	161	30	13
Kerala	14	14	321	6898	12263	1961	967	994	9034	4555	4479	1268	553	715	65	124
TamilNadu	29	29	796	16152	30733	4374	2226	2148	22727	11564	11163	3632	1892	1740	228	401
Pondicherry	2	2	44	977	1691	365	160	205	1063	541	522	263	131	132	10	12
Total	**584***	**567**	**16054**	**319239**	**720397**	**124856**	**66188**	**58668**	**527266**	**286374**	**240892**	**68275**	**38308**	**29967**	**4592**	**8763**

* Includes 8 districts of Mizoram and 2 of Andaman & Nicobar which were not surveyed this year

Sample Design of Rural ASER 2007

Dr. Wilima Wadhwa

The purpose of rural ASER 2007 is twofold: (i) to get reliable estimates of the status of children's schooling and basic learning (reading, writing and math ability) at the district level; and (ii) to measure the change in these basic learning and school statistics from last year. Every year a core set of questions regarding schooling status and basic learning levels remains the same. However a set of new questions are added for exploring different dimensions of schooling and learning in the elementary stage. The latter set of questions is different each year.

One government primary school in each sampled villages was visited during ASER 2005. This was not done in ASER 2006 but was done in ASER 2007. This year's ASER survey links the sampled child to the school that was visited. This will be able to better establish the links between schools and learning. Unlike, ASER 2006, this year mothers were not tested however, their education level was recorded. As in 2006, children were tested for comprehension. However, instead of testing only higher level readers, readers who are at a lower level were also tested. Finally, in ASER 2007 children were also tested for English reading and comprehension.

Since one of the goals of ASER 2006 Rural is to generate estimates of change in learning between 2006 and 2007, a panel survey design would provide more efficient estimates of the change. However, given the large sample size of the ASER surveys and cost considerations, we adopted a rotating panel of villages rather than children. In ASER 2006, we retained the 20 villages from 2005 and added 10 new villages. In ASER 2007 we randomly dropped 10 villages from ASER 2005, kept all the 10 villages from 2006 and added 10 more villages from the census village directory.

The sampling strategy used will generate a representative picture of each district. All rural districts will be surveyed. The estimates obtained will then be aggregated to the state and all-India levels.

Since estimates were to be generated at the district level, the minimum sample size calculations had to start at the district level. The sample size is determined by the following considerations:

- Incidence of what is being measured in the population. Since a survey of learning has never been done in India, the incidence of what we are trying to measure is unknown in the population.[1]
- Confidence level of estimates. The standard used is 95%.

Precision required on either side of the true value. The standard degree of accuracy most surveys employ is between 5 and 10 per cent. An absolute precision of 5 % along with a 95% confidence level implies that the estimates generated by the survey will be within 5 percentage points of the true values with a 95% probability. The precision can also be specified in relative terms — a relative precision of 5% means that the estimates will be within 5% of the true value. Relative precision requires higher sample sizes.

Sample size calculations can be done in various ways, depending on what assumptions are made about the underlying population. With a 50 % incidence, 95% confidence level and 5% absolute precision, the minimum sample size required in each strata[2] is 384.[3] This derivation assumes that the population proportion is normally distributed. On the other hand, a sample size of 384 would imply a relative precision of 10%. If we were to require a 5% relative precision, the sample size would increase to 1600.[4] Note that all the sample size calculations require estimating the incidence in the population. In our case, we can get an estimate of the incidence from ASER 2005. However, incidence

[1] For the rural sector we can use the estimates from ASER 2005 to get an idea of the incidence in the population.

[2] Stratification is discussed below.

[3] The sample size with absolute precision is given by $\frac{z^2pq}{d^2}$ where z is the standard normal deviate corresponding to 95% probability (=1.96), p is the incidence in the population (0.5), q=(1-p) and d is the degree of precision required (0.05).

[4] The sample size with relative precision is given by $\frac{z^2q}{r^2p}$ where *z* is the standard normal deviate corresponding to 95% probability (=1.96), p is the incidence in the population (0.5), q=(1-p) and r is the degree of relative precision required (0.1).

varies across different indicators --- so incidence of reading ability is different from incidence of dropouts. In addition, we often want to measure things that are not binary for which we need more observations.

Given these considerations, the sample size was decided to be 600 households in each district.[5] In each district, we have 10 villages from ASER 2005 and ASER 2006 and an additional 10 villages have been added this year to the sample, giving us a total of 30 villages per district. In each village 20 households are surveyed as in ASER 2006, giving a household sample size of 600 per district. National estimates from ASER 2006 put the proportion of children who could subtract or do more at 58%. If we use this as a measure of incidence, then our sample size of 600 would imply a relative precision of about 7% and an absolute precision of 4% at the district level to measure the proportion of children who could subtract. Note that at the state level and at the all-India level the survey has many more observations lending estimates at those levels much higher levels of precision.

If we had houselists at the district level, the 600 households could be randomly selected. In the absence of these, a two-stage sample design was adopted. In the first stage, 30 villages were randomly selected using the village directory of the 2001 census as the sample frame.[6] In the second stage 20 households were randomly selected in each of the 30 selected villages in the first stage.

Villages were selected using the probability proportional to size (PPS) sampling method. This method allows villages with larger populations to have a higher chance of being selected in the sample. It is most useful when the sampling units vary considerably in size because it assures that those in larger sites have the same probability of getting into the sample as those in smaller sites, and vice verse.[7, 8]

In the selected villages, 20 households are surveyed. Ideally, a complete houselist of the selected village should have been made and 20 households selected randomly from it. However, given time and resource constraints a procedure for selecting households was adopted that preserved randomness as much as possible. The field investigators were asked to divide the village into four parts. This was done because villages often consist of hamlets and a procedure that randomly selects households from some central location may miss out households on the periphery of the village. In each of the four parts, investigators were asked to start at a central location and pick every 5th household in a circular fashion till 5 households were selected. In each selected household, all children in the age group of 6-14 were tested.[9]

The survey provides estimates at the district, state and national levels. In order to aggregate estimates up from the district level households had to assigned weights --- also called inflation factors. The inflation factor corresponding to particular household denotes the number of households that the sampled household represents in the population. Given that 600 households are sampled in each district regardless of the size of the district, a household in a larger district will represent many more households and, therefore, have a larger weight associated with it than one in a sparsely populated district.

[5] Sample size calculations assume simple random sampling. However, simple random sampling is unlikely to be the method of choice in an actual field survey. Therefore, often a "design effect" is added to the sample size. A design effect of 2 would double the sample size. At the district level a 7% precision along with a 95% confidence level would imply a sample size of 196, giving us a design effect of approximately two.

[6] Of these 30 villages, 10 are from ASER 2005, 10 from ASER 2006 and 10 are newly selected in 2007. They were selected randomly from the same sample frame. The 10 new villages are picked as an independent sample.

[7] Probability proportional to size (PPS) is a sampling technique in which the probability of selecting a sampling unit (village, in our case) is proportional to the size of its population. The method works as follows: First, the cumulative population by village calculated. Second, the total household population of the district is divided by the number of sampling units (villages) to get the sampling interval (SI). Third, a random number between 1 and the SI is chosen. This is referred to as the random start (RS). The RS denotes the site of the first village to be selected from the cumulated population. Fourth, the following series of numbers is formed: RS; RS+SI; RS+2SI; RS+3SI; …. The villages selected are those for which the cumulative population, contains the numbers in the series.

[8] Most large household surveys in India, like the National Sample Survey and the National Family Health Survey also use this two stage design and use PPS to select villages in the first stage.

[9] In larger villages, the investigators increased the interval according to a rough estimate of the number of households in each part. For instance, if a village had 2000 households, each part in the village would have roughly 500 households. Selecting every 5th household would leave out a large chunk of the village un-surveyed. In such situations, investigators were asked to increase the interval between selected households.

The advantage of using PPS sampling is that the sample is self weighting at the district level. In other words, in each district the weight assigned to each of the sampled household turns out to be the same. This is because, the inflation factor associated with a household is simply the inverse of the probability of it being selected into the sample times the number of households in the sample. Since PPS sampling ensures that all households have an equal chance of being selected at the district level, the weights associated with households in the same district are the same. Therefore, weighted estimates are exactly the same as the un-weighted estimates at the district level. However, to get estimates at the state and national levels, weighted estimates are needed since states have a different number of districts and districts vary by population.

Even though the purpose of the survey is to estimate learning levels among children, the household was chosen as the second stage sampling unit. This has a number of advantages. First, children are tested at home rather than in school, allowing all children to be tested rather than just those in school. Further, testing children in school might create a bias since teachers may encourage testing the brighter children in class. Second, a household sample will generate an age distribution of children which can be cross-checked with other data sources, like the census and the NSS. Third, a household sample makes calculation of the inflation factors easier since the population of children is no longer needed.

Often household surveys are stratified on various parameters of interest. The reason for stratification is to get enough observations on entities that have the characteristic that is being studied. For instance, the NSS uses a two stage stratified sample for their consumption surveys. In the first stage the sample is stratified by population and in the second stage households are stratified on the basis of their affluence. The reason for doing this is that the purpose of the survey is to generate poverty estimates for which a representative sample must include enough non-affluent households. The ASER survey stratifies the sample by population in the first stage. No stratification was done at the second stage. Since the proportion of population in the 6-14 age group is about 22% and the average household size is about 5,[10] a simple random sample at the second stage would yield enough children in the sample. Finally, if we were to stratify on households with children in the 6-14 age group, we would need the population of such households in the village, which is not possible without a complete houselist of the village.

[10] NSS 55th Round.