Sample note

Cambodia Public Expenditure Tracking Survey

April 2004

The World Bank

The objective of the study is twofold; to conduct a public expenditure tracking survey in primary education, and to assess the impact of increased funding to primary schools on their learning outcomes. In meeting the second objective, the study exploits the fact that increased funding (sent directly to the schools) was phased-in over time. The sample will also be linked to an existing household survey to allow answering several other related research questions.

The sample consists of 220 primary schools randomly drawn from 12 provinces and 34 districts. The sampling procedure involves two main stages. The first stage entails drawing provinces and districts from the 2002-03 school census, while the second stage narrows down the eligible schools to the ones situated in communes where the Cambodian 2003-04 household survey is currently being carried out. The second stage implies both potential benefits and costs. In terms of benefits, it provides important background variables for the schools to be visited. But at the same time it limits the pool (the final pool of schools corresponded to 45% of the total population of primary schools in the country) from which the final schools are drawn from. The latter would be a problem if there were difficulties, such as the sample not being representative or there existed systematic biases in terms of the schools finally included. As will be discussed in more detail below these issues should not be a problem, however.

The first stage of the sampling is based on the 2002-03 school census. Three general criteria governed the choice of procedure in selecting the sample in the first stage. First, the sample would be representative of the population of early provinces (receiving funding from 2000 and onwards) and late provinces (receiving funding from 2001 and onwards) in the country. Second, the sample of provinces chosen would also be representative in terms of the number of schools in general and number of early and late schools in specific. Third, the districts picked within each chosen province would have to balance the aim of being representative in terms of the number of schools in the districts, and at the same time represent a manageable task for the survey team in terms of accessibility. To account for these three considerations, a stratified random sample was chosen. Specifically, each province was weighted according to size (number of schools). Thereafter, 5 early and 7 late

provinces were randomly chosen from the population of 24 provinces. Each of the 12 selected provinces (Banteay Meanchey, Battambang, Kampong Cham, Kampong Chhnang, Kampong Speu, Kampong Thom, Kampot, Phnom Penh, Ratanak Kiri, Siemreap, Svay Rieng, and Takeo) was then allotted a number of schools based on the proportion of schools in the province to the total number of schools in the selected sample of 12 provinces. To make the survey effort is feasible, it was decided that 2 or 3 districts would be picked from each province (the final number depending on the total number of districts in the province). Similar to the draw of provinces, each district was weighted according to size (number of schools). Thereafter, 2 or 3 districts were randomly chosen from the population of districts within each province, yielding a total of 34 districts to be included in the final sample. See Table 1 and Table 2 below for details.

The second stage of the sampling is based on the Cambodian 2003-04 household survey (HSES). The HSES sampling frame consists of 900 villages and 15,000 households. The sampling design involved stratification of the country into five geographical regions (Phnom Penh, Plain, Tonle Sap, Coastal, and Plateau and Mountain), dividing up each region into separate urban and rural strata. From each stratum, 4 independent sub-samples of villages were drawn, with the sample being allocated over the strata proportionately to the total number of households in the strata (see Statistics Sweden, 2003 and World Bank, ???? for details). Two considerations played a role when matching the HSES with the 2002-03 school census. First, as the main survey unit for the HSES was the village level, ideally one would like to match as many schools as possible from the census with ones residing in a village where the household survey had been carried out. In theory, this would yield a total of 900 eligible schools to make the final draw from – *if* all villages surveyed had a primary school. This turned out to be problematic for two reasons. First, even if all 900 villages in the HSES did have a primary school this is only accounted for 15 percent of the total population of 5,915 schools in the country. Second, in reality there were only about 450 villages included in the HSES that did have a primary school. Hence, to broaden the number of eligible schools, the HSES and the census were matched at the commune, rather than the village level. In other words, schools were included from the census if they were situated in a commune where the HSES had been carried out. This procedure reduced the number of eligible primary schools from 5,915 to

2,689. The second concern in the matching was that the sample of urban and rural schools would be representative of the population of urban and rural villages included in the HSES. To meet this condition, the number of schools in each district arrived at in the first stage was weighted into an urban and a rural group. The groups were weighted according to the proportion of urban and rural villages that the HSES ascribed to the specific province which the district belonged to. Finally, schools were randomly drawn from the group of 2,689 schools, with the number of schools allotted to each district being decided in the first stage, while the urban-rural weight followed the HSES. The final draw resulted in a slight overrepresentation of rural schools as compared to the share ascribed by the HSES; see Table 3 below for details.

Matching the HSES with the school census may be problematic for two reasons. First, if the villages included in the HSES did not present a representative sample of the population of villages in the country, say by under representing urban villages, the match would suffer from the same under representation. There is no reason to believe that this is the case however, as the sampling procedures governing the finalization of the HSES seems to have been statistically rigorous in this regard (yielding a representative sample of the country as a whole). Second, despite the HSES being representative, the match may still suffer from a bias if there is some systematic reason as to why certain schools were included in set of eligible schools from which the final draw was conducted. Again, this should not be a problem. In fact, comparing summary statistics of some key variables of the 2,689 schools included with those excluded, shows the two groups are fairly similar, with a slight over representation of urban and larger schools, see Table 4 below (note that here we are comparing the census with the eligible group from which the final draw was made, not the HSES as referred to in Table 3).

Table 1

	Early provinces	**Late provinces**
Percentage of provinces sampled	42	58
Population percentage of a specific province	42	58

Table 2

Early provinces	**Number of schools**	**Number of districts**	**Late provinces**	**Number of schools**	**Number of districts**
Kampong Chhnang	19	3	Banteay Meanchey	19	3
Kampong Speu	19	3	Battambang	19	3
Kampot	20	3	Kampong Cham	25	3
Ratanak Kiri	19	3	Kampong Thom	20	3
Svay Rieng	19	2	Phnom Penh	10	2
			Siemreap	20	3
			Takeo	20	3
Total	**87**	**14**	**Total**	**133**	**20**
Percentage of early schools to total schools sampled	40		**Percentage of late schools to total schools sampled**	60	
Population percentage of early schools to total population	39		**Population percentage of late schools to total population**	61	

Table 3

	Planned Number of Urban Schools	**Actual Number of Urban Schools**	**Planned Number of Rural schools**	**Actual Number of Rural schools**	**Total (Actual)**
Banteay Meanchey	7	7	12	12	19
Battambang	7	6	12	13	19
Kampong Cham	2	0	23	25	25
Kampong Chhnang	4	0	15	19	19
Kampong Speu	3	0	16	19	19
Kampong Thom	6	6	14	14	20
Kampot	3	0	17	20	20
Phnom Penh	8	10	2	0	10
Ratanak Kiri	4	4	6	6	10
Siemreap	7	6	13	14	20
Svay Rieng	2	0	17	19	19
Takeo	3	0	17	20	20
Total	56	39	164	181	220

Table 4

	Population	**Included schools**	**Excluded schools**
Total number of schools	5915	2689	3226
Number of students on average	457	536	404
Number of teachers on average	8	10	7
Percentage of urban schools	11	15	7
Percentage of rural and remote schools	89	85	93