



ISCWEB – Children's Worlds Jacobs Foundation project

A NOTE ON SAMPLING

Jonathan Bradshaw & Gwyther Rees

The sample for the main stage is supposed to be representative of a country (ideally), region or some other coherent population area.

For a sample to be representative each unit of analysis has to have an equal chance of selection. In this project the unit of analysis will be the child but the primary sampling unit will be the school. We are aspiring to achieve a representative sample of at least 1000 eight year olds, 1000 ten year olds and 1000 12 year olds.

In order to achieve this we need to start by drawing a sample of schools which will achieve this number. The number of schools we need will depend on their size and how many children we plan to survey in each school. As described below, we would expect to sample at least one class group per school. So, for example, if the average class group in the sample is 25 children we would need a total sample of around 40 schools for each year group. In some circumstances as discussed below it may make sense to sample more children than this in some schools, but this should be done with caution as reducing the number of schools increases the clustering within the sample.

So we need to start with a list of schools. It would be extremely helpful if we stratified our sample of schools at this stage to ensure that important sub-groups within the sample are adequately represented. There are a number of ways which we might be able to stratify schools. By:

- Region
- Type
 - Primary/secondary
 - Private/public
 - Religious affiliation
 - Rural/urban
 - Size
 - Gender

Doing this kind of stratification will depend on how much information is available in advance about schools.

We probably won't all be able to stratify by all these factors. But the aim is to select a sample of schools that will generate a representative sample of children of the right numbers and ages. Schools should be selected by random from the list and then if you find you cannot get access to the school it should be replaced by another school with the most similar characteristics or randomly if you don't know these.



One important issue to consider in drawing the sample is how much variation there is in school sizes within the country or region. If there are substantial variations in school sizes and the sample draws a fixed number of pupils from each school (e.g. a class group) then children in smaller schools will have a higher probability of selection than children in larger schools. This introduces a potential bias if there is a connection between school size and other factors - e.g. smaller schools might tend to be in rural areas or in more or less affluent areas. There are two possible solutions to this problem:

1. If information on the number of pupils in every school in the sampling frame is available when the sample is being selected, then schools can be initially selected with probabilities proportional to size. (Further guidance on this technique can be provided if needed). Then if equal numbers of children are sampled in each selected school, each child in the sample will have an equal probability of selection.
2. If (1) is not possible, then it can still be helpful to gather information on school size from each of the schools selected. Then, for example, the selected schools could be divided into two groups (large and small) and twice as many children (e.g. two class groups) could be surveyed in the larger schools as in the smaller schools. This will improve the overall balance of the sample.

The next stage is to sample students within schools. The key here is to minimise the chances of a bias in the selection of students. The most likely source of bias is if there is class streaming. We need to avoid this. In England we managed to do this by choosing mixed ability year group classes – used for some lessons. If there are no mixed ability classes then it will be necessary to sample students across classes to avoid ability bias. The key here is that the students are selected randomly preferably from a list. In mixed gender schools a representative balance of genders is essential. If selected children are not present on the day then effort needs to be made to get responses from these children when they do attend.

It is important to keep a record of the sampling strategy. Record response at the school sampling stage (in our experience non response is a serious problem at this stage) and at the student selection stage (in our experience non response is not a problems at this stage). Bias is most likely to arise from school selection and non response. For example in England we ended up with too many private school children in one of our surveys.

A sample will never match the population exactly. This is because of sampling error and it is respectable and predictable. It can be minimised by random selection, stratification and increasing sample size. Bias is not respectable. However we can adjust post hoc for known bias. So for example if we know that our sample of schools and therefore students is biased in favour of private schools we can adjust for this by reweighting the sample at the analysis stage. But we can only do this if we know that there is bias in the selection. So it is essential to compare the achieved samples with the selected samples in terms of key characteristics. However, weighting can also create other problems, particularly in

CHILDREN'S
WORLDS



INTERNATIONAL SURVEY OF CHILDREN'S WELL-BEING (ISCWeB)

JACOB'S
FOUNDATION

Our Promise to Youth

small samples, where certain cases may be highly weighted and could have a strong influence on findings. So, the first priority should usually¹ be to gain the most representative sample possible and weighting should then be used to correct for any problems that arise.

¹ Note: One other possible use of weighting is where you wish to ensure that you gain an adequate sample of a particular minority sub-group for comparative analysis. In this case, it is possible using stratification to include a larger sample of this sub-group than is representative. Then weighting can be used to reduce this sub-group to a representative proportion when overall statistics are being calculated. However, if you wish to do this in the Children's Worlds survey then it will be important that the final weighted sample size is at least 1,000 children.