FAQ 10: How do you sample children for quantitative research?

What's the issue?

Sampling for quantitative research depends on whether or not researchers are aiming for a probabilistic sample from which they would like to draw inferences about the population (i.e., to what extent sample statistics reflect the population parameters). A number of issues usually have to be considered (choosing the population, the sampling frame, the way of sampling, and the sample size). When aiming for a representative sample, things get more complicated as researchers need a list of children to sample from. This can be quite difficult. However, they could establish a list of all schools in the area they are interested in studying, and sample children from there.

Common practice

- When conducting a survey with children and parents, the household could be used as unit of analysis.
- If financial or time sources do not permit face-to-face surveying at home, children could be sampled by school (e.g. a sample based on clusters), covering different regions of the country. Instead of individual children, sample groups of children occurring naturally in the population could be sampled – this is known as cluster sampling.
- If researchers want various subgroups (e.g. age subgroups or gender subgroups of children) in the sample to also be representative, stratified random sampling can be used, which combines stratified sampling with random sampling. For example, if researchers wanted a stratified random sample of boys and girls from the final year of a primary school, they would first separate the entire population of the last year of the primary school pupils into two groups, one all boys and the other all girls. To complete the sampling they would then independently select a random sample from each stratum (a random sample of boys and another of girls).
- Researchers can also do a non-probabilistic sample of children, bearing in mind that no inferences beyond their sample are possible. However, studies with non-probabilistic samples (e.g. quota sample, purposive sample) are still valuable as they can be very informative, and also point to the children beyond the sample who most probably have very similar socio-demographic characteristics to those included in the sample. It is okay to conduct such studies as long as the aim is not for statistical inferences from the samples to the population. The researchers operate only within descriptive interpretations.

Questions to consider

What size should the sample be? Do you need probabilistic sampling? Can you afford to sample probabilistically? What kind of natural clusters of children are available in the population? Do you also need various subgroups in the sample to be representative?

Pitfalls to avoid

- Sampling children requires consent from the parent or legal guardian, as children are underage and cannot commit to such decisions on their own.
- Sampling requires that all parameters of concern are taken into consideration (e.g. socio-economic background, gender, etc.). Missing one parameter of interest, or not fully representing it, could bias the results.
- Time needs to be taken to ensure that the sample is well defined and complete (e.g. a sample of children who only speak English does not represent all children in the country, as a population might include minority ethnic children who do not speak English).

A researcher's experience

> In designing a national survey for children, as it was too expensive to interview children in their households, it was decided to sample children by schools. This sample was based on clusters covering different regions of the country. After negotiation with the Portuguese Minister of Education, it was agreed that in each of five regions, four elementary schools attended by children (6–15 years) would be selected based on the criteria of urban/rural contexts, children from minority ethnic groups and socio-economic

status (SES). Based on lists of students in each of the 20 schools, a proportional sample of children by age would be designed and 30 students from each school were then chosen randomly. This way, the sampling would involve 600 students.

After parental consent had been obtained, the sample would receive a self-completion questionnaire to be answered at school, outside the classroom. Parents would receive another self-completion questionnaire, given to them by the child in the study. These questionnaires were to be returned to the school, in closed envelopes, and the school would send them to the research project.

This initial design for a national survey proved to be too difficult and time-consuming. It involved several factors, starting with the agreement of the schools randomly selected, and ending with the parental consent of all the students randomly sampled.

Instead, it was decided to sample children by schools in the greater Lisbon area, which is the leading area for internet penetration in households and the area with more migrant children. The Minister of Education provided us with a list of the public elementary schools covering compulsory education in this area. From this list, 20 schools were selected based on the criteria of urban/rural contexts, children from minority ethnic groups and SES. The first 11 schools that accepted the idea were our sample. Each school chose a class per year from the 4th to the 8th grade, providing an average of 90 children as a starting point. Parents were asked for informed consent. In each school, children who had parental consent were presented with the aims of the research and invited to participate, under the assurance of privacy and confidentiality. The self-completion questionnaire was answered at school, in the presence of an assistant, a member of the research team. Parents received another self-completion questionnaire, given to them by the child in the study. These questionnaires were returned to the school, in closed envelopes. In order to provide identification, children and parents' questionnaires had the same code number.

In the end, a total of 810 questionnaires answered by children at school and 630 questionnaires answered by their parents were sent to the research team, which might be considered quite a positive number. Parents who answered this questionnaire differed from the national profile – they were better informed and had higher levels of education. Also, parents of younger children (9–11) were overrepresented compared to the parents of older ones (12–14), and this may have different meanings, including the possibility that the older children may have resisted involving their parents. (Cristina Ponte, Portugal)

In the TIRO research project we organized two panels of 20 Dutch- and 20 French-speaking teenagers (aged 12–18). We interviewed them, had online conversations with them on several occasions, and asked them to keep a diary on their everyday life and media use. For sampling those panels we went to different sites where young people are present (schools, youth movements such as the scouts, and youth clubs, e.g. for sport, theatre), and we used our own social networks, although no close relatives were selected, only casual acquaintances. In order to manage the subjectivity in the sampling process (two researchers were involved and we wanted to avoid discrepancy between the Flemish and Walloon panel), we used a theoretical sampling matrix. First, the hundreds of young people we recruited were asked to provide brief information about their social background, ICT use, and leisure. Based on a literature review we then decided to sample both panels by means of three criteria that seemed to be distinct for explaining the diversity and heterogeneity of young people's internet practices: gender, age (12–13, 14–16, 17–18), and SES (reflecting the economic and cultural capital of the parents). Based on these three socio-demographic characteristics we drew a matrix with 18 cells and looked for young people who met the cell criteria that were preconceived (e.g. one boy aged 12–13 with low SES, one girl aged 14–16 with high SES). To gain insight into future trends in ICT use, we also selected in each panel one teenager who showed an intensive pattern of ICT use. This sampling procedure (in stages and pre-structured) proved to be useful in guaranteeing the diversity of the panel. We wanted especially to avoid assembling a middle-class panel, since many qualitative studies seem to suffer from this bias. Yet we did not succeed in involving young people with a minority ethnic background in our panel. More specific sampling methods seem to be required for including those groups. (Joke Bauwens, Belgium)

References and further resources

Kalton, G. (1983). *Introduction to survey sampling*. Newbury Park, CA: Sage Publications.

Sapsford, R. (2007). *Survey research*. London: Sage Publications.