



ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT

**HIGHER EDUCATION INSTITUTIONS:
SAMPLING ISSUES PREPARING THE AHELO FEASIBILITY STUDY**

**Jean Dumais, Chief, Statistical Consultation Group,
Statistics Canada, 3 December 2008.**

Presented at the OECD AHELO GNE Meeting in Paris on 17-18 December 2008

1. Introduction: Sampling for a Feasibility Study	2
2. Random or Purposive Sampling of Institutions	3
3. Sampling of Students	3
3.1 The Target Population and the Survey Population	3
3.2 The Sampling Frame	4
3.3 Preparing the Sampling Frame for Sample Selection.....	4
3.4 Selecting the Student Sample	5
4. Response Rates and Standards.....	6
5. Response Incentives.....	7
6. Ask Your Experts.....	8
REFERENCES	9

1. Introduction: Sampling for a Feasibility Study

In preparing any large-scale survey, it is desirable – and customary – to do a feasibility study. Typically, a feasibility study will help in finding out what survey step works well, what step needs improvement and, possibly, what step deserves to be redesigned.

In principle, a feasibility study can be tailored to address specific issues, for example:

- Determine whether the concepts and constructs identified by the subject matter specialists can be measured;
- Determine how to best apportion the workloads to the national, regional and local organisations;
- Obtain advanced data to help in the creation and testing of data models;
- Test and/or validate the survey instruments, questionnaires and forms;
- Test and/or validate the various means to create a sampling frame, and assess its quality;
- Test and/or validate the various sampling procedures and forms;
- Test and/or validate the collection procedures, document routing and interviewer feedback protocols;
- Determine the level of human and physical resources that the full operation would likely require;
- Assess the effectiveness of public relations materials and campaign;
- Determine whether response is likely to be adequate;
- Identify areas of higher non-response where field effort might require additional investment.

In some situations, especially when the survey of interest is a full or partial repeat of an earlier instance, only selected steps may require testing; for ground-breaking projects, it is often safer and wiser to test as broad a range of steps as possible; it is also more expensive.

The OECD has elected to conduct a feasibility study for the Assessment of Higher Education Learning Outcomes (AHELO) before it be launched on a full-size scale. The OECD has set two main goals for the AHELO feasibility study, namely:

- “Test the science of the assessment – whether it is possible to devise an assessment of the outcomes of tertiary education which enables to make reliable statements about the performance/ effectiveness of learning in [HEI] of very different types, and in countries with different cultures and languages; and
- Test the practicality of implementation and of motivating [HEI] and students to take part in such a test.” (OECD, 2008, paragraph 22).

In this brief paper, we will review some aspects of sampling as they pertain to surveying and education assessment in Higher Education Institutions (HEI). We will also touch on some collateral issues.

2. Random or Purposive Sampling of Institutions

Depending on the goals of the feasibility study, life-like conditions may be desirable; in that case, the project management team will try and ensure that the conditions under which the survey is likely to be conducted are reproduced as faithfully as possible during the feasibility study. This would apply whether the full-size study be a national all-encompassing survey to feed into an international comparative study or be limited in scope for political, scientific or organisational reasons. Alternatively, some aspects of the main study may not require thorough prior testing and the feasibility study may be limited to selected aspects. Then, life-like conditions are not required and need not be duplicated.

Under the first scenario where life-like conditions should be reproduced, it is likely that random sampling would be applied at every stage of the feasibility study, provided that budgets, infrastructure, time and human resources are available in sufficient amounts. This strategy also assumes that the project management team accept that all situations might not materialise during the feasibility study and, therefore, that some areas remain grey.

Random sampling of institutions would typically allow the project management team to get a sense of institutional response (and non-response), to assess the smoothness of the systems integration, and to set initial values for some parameters of the main study. Random sampling would also mean additional work for frame construction and validation, determination of an efficient sampling strategy and adequate sample size, the loss of control over some of the characteristics of the participating institutions and no guarantee of a useful response rate.

Under the second scenario, purposive or convenience sampling may suffice, assuming that the project management team understand and accept the limitations of their sample. Clearly, advanced estimates of institutional non-response cannot be obtained, but might not be relevant. Working with volunteer institutions eludes frame preparation and validation, almost guarantees response. While the project management team, along with the national project manager (NPM), can almost preset the characteristics of the participating institutions, the scope of the lessons learned is also limited to those characteristics.

Whereas PISA is an international comparative study of national outcomes, AHELO is not meant to be such a study (*ibid.*, paragraph 17); thus, there is no requirement to obtain and use random samples of institutions. AHELO project management team has determined that a sample of ten volunteer institutions per participating country would suffice.

3. Sampling of Students

3.1 *The Target Population and the Survey Population*

The project management team, or a designated group of subject matter experts, will determine what students are targeted by the AHELO, that is those about whom any inference or conclusion is to be made. The target population is often a theoretical statement more than a practical definition of who to survey. The current approximation of the target population is “*towards the end of a three or four-year degree*” (OECD, 2008, p. 6), together with the requirement of assessing learning outcomes from a *generic skills strand*, a *discipline strand*, a *value-added measurement strand* and a *contextual strand*. The last two strands are perhaps more relevant to the institution itself than to its students.

These study objectives will be (have been?) translated into something more operational, indicating what students of what programmes, departments or faculties should be considered in-scope. Even with this refinement, the actualisation of the target population in each institution may be difficult. Moreover, because of operational constraints, both local and national, some of the targeted students may need to be excluded from the study, loosening the connection between the theoretical and practical views of the

population of interest. Defining the survey population of each institution and ensuring that it is comparable in intent, composition and coverage to that of the other participating institutions within a country and across participating countries will prove to be a challenge.

In essence, if the target population comprises those one wishes to talk about, the survey population comprises those one can realistically talk about.

3.2 The Sampling Frame

In most of the participating countries, it would likely be easy to compile a complete and up-to-date list of HEI. However, in all likelihood, a comprehensive, national, complete, up-to-date and unduplicated list of undergraduate students will not exist in most participating countries, let alone be available to the NPMs or to their institutional representatives. Hence, one of the key steps of feasibility study will be the creation of a high quality student frame, within each participating institution.

The main characteristics of a high quality sampling frame are (Statistics Canada, 2003, p. 28):

- **Relevance:** the extent to which the frame corresponds to and allows access to the target population;
- **Accuracy:** the extent to which the list is complete, exhaustive, free of out-of-scope units, unduplicated; the extent to which classification variables are available and correct for each entry;
- **Timeliness:** the time difference between the moment the frame is created and the moment of data collection;
- **Cost:** the cost of creating and obtaining the frame and the cost of up-dates to the original listing if timeliness is dubious.

The first challenge is thus to compile an exhaustive, exclusive and up-to-date list of (say) “students in the final year of any undergraduate programme.”¹ It may already exist, as an institutional register; but experience shows that this need not be true in every institution of the same country. One has to suppose, here, that the institutions, having agreed to participate in the feasibility study, will make their lists available to the NPM or their institutional representatives.

3.3 Preparing the Sampling Frame for Sample Selection

Economics and engineering having been identified as the focus of the discipline strand, stratification of the student frame by discipline appears natural. Three strata may be sufficient: those students enrolled in an engineering undergraduate programme, those enrolled in an undergraduate economics programme, and the rest of the students. Stratification creates independent (sub)-populations from which independent samples can be drawn. More strata could be created, but at least those three would seem required.

The sampling frame should contain, for each student, a unique identifier (preferably the name), how or where to reach the student, and to what stratum he or she belongs; other data fields may prove to be useful: courses currently taken, previous degrees, number of credits or courses successfully taken, name and type of programme, gender, age or date of birth, birthplace or permanent residence, etc. Some of these extra variables may prove useful at the sample selection stage, while others may be useful at the later

¹ To lighten the remainder of this document, “student” refers to one enrolled in the final year of any undergraduate programme, unless specified otherwise.

stages of data cleaning and estimation. If a variable is thought to be useful, it should be available for each student on the frame.

Problems of defining the strata may occur when, say, an undergraduate degree in economics may not exist per se, but some combination of courses within a recognised programme may amount to the same. These problems will need to be resolved institution by institution while agreeing with the standards and objectives of the feasibility study, most likely with the opinion or advice of the AHELO project management team.

3.4 Selecting the Student Sample

Depending on the size of the stratum (how many students are enrolled) and on the required sample size per stratum, it may or may not be necessary to sample; it is often less work to survey every member of a stratum when the sampling fraction, f_h , is between 0.75 and 1.0 ($f_h = n_h \div N_h$, that is, the required sample size divided by the stratum size). In such situations, sampling would not be justified and all members of the stratum would be invited to participate in the study.

In strata where the sample size is small enough compared to the stratum size, then a sampling strategy must be determined for each stratum and each institution. It may be a unique strategy applicable to all institutions and all strata, or it can be adjusted to the local conditions, which is more work, but often preferable. The only rule is that any sampling strategy must ensure that the samples be unbiased and random so that proper design-based estimates of student characteristics and scores be obtained as well as proper estimates of sampling error.

Several classes of sampling strategies are available: unit sampling and cluster sampling, equal probability or unequal probability sampling, simple or systematic sampling, with or without replacement sampling. These can be combined in more or less efficient or practical strategies, for example, “unequal probability sampling of units without replacement” or “systematic equal probability sampling of clusters without replacement”². The comparative statistical advantages of those strategies have been amply documented in the literature to which we refer the interested reader (Kish, 1965; Lohr, 1999; Statistics Canada, 2003).

3.4.1 Unit Sampling

If unit (*i.e.* student) sampling is preferred or more practical, then the sampling frame must comprise every in-scope student, and only them, once and only once, and each student must be assigned to one and only one stratum. Simple algorithms exist to draw simple random and systematic random samples of fixed determined size (Tillé, 2001).

Sampling routines offered by general purpose software for general office work (*e.g.*, spreadsheet applications) should be used with extreme caution as these routines are often limited, flawed or poorly documented.

3.4.2 Cluster Sampling

In some institutions, it may be required by its administration that the students be sampled by entire classrooms, rather than individually from a list. This is possible, of course, but a few elements of planning must be kept in mind. If sampling is to be done by classroom (or some equivalent grouping of students), then the institutional representative must make sure that every student is indeed assigned to one of the

² By definition, systematic sampling can only be done without replacement.

classes and that these classes do not overlap. For example, all engineering students may be required to take some given mathematics or foreign language class, and they all take it the same term, without exception. Furthermore, those classes should only contain one type of students, *e.g.*, engineering.

Then, simple or systematic sampling, with equal probability or with probability proportional to size, can be used for the selection of (ideally) two classes per stratum per institution.

If cluster sampling is implemented in one institution, it need not be implemented in all institutions (though a unique strategy is often simpler to manage centrally).

3.4.3 Comprehensive sample vs. strand-specific

Two main sets of student skills are of interest to AHELO: cross-sectional comprehensive skills and discipline-specific skills. The overall study strategy should determine if (part of) the cross-sectional skills will be tested in the discipline-specific instruments or if some of the students selected for the discipline strand will be used to measure some of the cross-sectional skills rather than discipline-specific skills. These options will dictate to some extent how the sample of students can be selected.

3.4.4 Sample size and expected precision

Organisational constraints (booklet rotation patterns, budgets and calendar, human resources, collection calendar, etc.) will dictate to a large extent how many students are expected to be sampled for AHELO. The expected response rate also plays a role in that determination. If funds and prior information on the scales of interest are available, some consideration could be given to the expected level of precision (sampling error). The size of the institution, of the cohort or of the classes will also be a factor.

The project management team should determine what level of precision is required of each strand, each scale, each country, if applicable.

3.4.5 Summary and key factors

Unless, the institution is quite small, random sampling of students is quasi-compulsory if any statistically meaningful institution-level statistics are to be obtained. Random sampling will also give an early estimate of what the student response is likely to be during the main study.

The project management team and the various national project management teams should keep in mind that (1) the national structure of higher education; (2) the structure of the selected institutions; (3) the institutional Ethics Committees; (4) the overlap of instrument contents; (5) the timing of the study in the academic year; (6) the availability of an institutional frame; (7) local requirements or contingencies; (8) the separation of responsibilities between the national centre and the institutional representatives, all play a key role in determining an effective national strategy for the sampling of students.

Random sampling of students should be performed under some serious amount of guidance or supervision: reproducibility of the samples is, to a large extent, a guarantee of unbiasedness (that is also why spreadsheets are not a good idea) and of credibility.

4. Response Rates and Standards

There is but one international standard for response: the higher the better!

Still, most, if not all, educational assessments have established what they consider to be the minimum response before the project team asks for some form of analysis of the (non-) response bias.

For example, PISA 2003 sets its target response rates at 85% of the initially selected schools, and 80% of the selected target students in the participating schools (OECD, 2005). Those requirements were set based on experience with the previous waves of PISA and with an aim at minimising response bias and sampling error (*ibid.*, pp. 48-49). Results for PISA 2003 show school response rates between 64% and 100%, and student response rates ranging from 77% to 98%, depending on the country (*ibid.*, pp. 171-173).

For their part, TALIS, a survey of active lower secondary school teachers, have set their target response rates at 75% of the schools and 75% of the students in participating schools (OECD, 2007, pp. 16-17). This is slightly lower than what is required for PISA; the project management team thought that, as one of the first overall surveys of ISCED Level 2 teachers, TALIS could achieve those targets. Results are compiled but currently under embargo.

Lastly, the Teacher Education and Development Study – Mathematics (TEDS-M), sponsored by the IEA, adhered to the IEA standard participation requirements of 85% for both institutions and individuals (IEA, 2007). This survey, of the three listed here, is the one whose target population is most comparable with that of AHELO, namely, students of Teachers' education programmes in their final year of schooling. Data on participation are compiled but results are currently under embargo.

5. Response Incentives

The recent TEDS-M has highlighted a number of incentives that can help increase response. Preliminary indicators would tend to show that securing the participation of an institution is far easier than securing that of undergraduate students, or of faculty members, at the end of their school year. The National Research Co-ordinators of TEDS-M have highlighted some of the actions that they have undertaken in their respective countries that have proven valuable incentives:

- Lots of public relations and communications; support letter from the minister of higher education; support letter from president or principal of the institution, from the various department heads to their respective students, announcements in scientific journals, advertisements in professional associations' newsletters...
- Establishing a local on-line newsletter
- Recruiting institutions, institutional co-ordinator very early in the process;
- Organising seminars announcing and explaining AHELO and its benefits to the community;
- Monetary gift to the participating institutions, departments or local co-ordinators (bookstore vouchers, special grants, ...);
- Lotteries (bookstore vouchers, "iPod", ...), small gifts (pens, mugs...) to participants;
- Honorarium;
- Others, as applicable or as possible.

It appears that in many countries and in many institutions, response was better when the co-ordinator was independent from the institution.

Problem resolution during collection is a response-maintaining, rather than -inducing, tool: having no one to turn to in case something goes wrong may make some co-ordinators feel unsupported, unheard by their national centre.

Now, one may ask: "do incentives augment response bias"?

6. Ask Your Experts

At this point, this early in the lifecycle of the project, there are several areas still in several tones and shades of grey. With time, one can expect that their picture will become clear and sharply in focus.

What follows is a short list of topical questions that have some bearing on sampling of institutions and of students. They don't all need immediate answers, but they will need an answer one day. Those questions are grouped according to who might be in a better position to provide an adequate and working answer.

- *Subject Matter:*
 - Who is in scope?
 - Who is not?
 - What to do with individual non-response?

- *Project Management Team:*
 - Is confidentiality guaranteed? How? By whom?
 - Is anonymity of responses guaranteed? How? By whom?
 - How are institutions "protected"?
 - Are there options for the mode of collection?
 - Who will have access to micro-data?
 - What is the collection window? Is it movable, stretchable?
 - How are identical testing conditions secured?
 - What quality indicators are contemplated?
 - What are the minimum participation rates?

- *Scaling:*
 - How many booklets will there be?
 - If more than one, how many of each are required for reliability of scales?
 - If more than one, are they independent "booklets" or partially overlapping "booklets"?
 - What student contributes to what score, scale, result, table?

- *Sampling:*
 - Is there a fixed desired level of precision? If so, at what level (national institutional, international, scale...)?
 - How many of each (booklets, students, institutions) are required for precision?
 - What to do with non-response of institutions, of individuals?
 - Are replacement units allowed?
 - How are the estimation weights and adjustments going to be computed?
 - What variance estimation method is expected to be used?

REFERENCES

IEA (2007), *TEDS Sample Preparation Manual*, TEDS-M Ref. No. MS-02-01, Amsterdam and Hamburg.

Kish, L. (1965), *Survey Sampling*, Wiley, New York.

Lohr, S. (1999), *Sampling: Design and Analysis*, Duxbury Press, Pacific Groves, CA.

OECD (2005), *PISA 2003 Technical Report*, Paris.

OECD (2007), *Teaching and Learning International Survey (TALIS) Sampling Manual*, TALIS Ref. No. MS-02-03, Paris.

OECD (2008), *Roadmap for the OECD Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study Roadmap*, <http://www.oecd.org/dataoecd/50/50/41061421.pdf>.

Statistics Canada (2003), *Survey Methods and Practices*, Catalogue Number 12-587-XPE, Ottawa.

Tillé, Y. (2001), *Théorie des Sondages, Échantillonnage et estimation en populations finies*, Dunod, Paris.