

**DIRECTORATE FOR EDUCATION  
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD**

**Group of National Experts on the AHELO Feasibility Study**

**SAMPLING MANUAL**

**8th meeting of the AHELO GNE**

**Paris, 28-29 November 2011**

*This document was prepared by the ACER Consortium and is annex 3 to the phase 2 progress report.*

*It is only available in PDF format.*

Contact:  
Consortium: [ahelo@acer.edu.au](mailto:ahelo@acer.edu.au)  
OECD Directorate for Education: [Diane.Lalancette@oecd.org](mailto:Diane.Lalancette@oecd.org)

**JT03311450**

Document complet disponible sur OLIS dans son format d'origine  
Complete document available on OLIS in its original format



EDU/IMHE/AHELO/GNE(2011)21/ANN3



## Sampling Manual

October 2011



Statistics  
Canada

Statistique  
Canada



## Table of contents

Introduction.....	3
About this AHELO Sampling Manual .....	3
AHELO’s goals .....	3
Sampling plan for AHELO .....	4
NPM responsibilities in sampling .....	5
Sampling Team responsibilities .....	5
Overview of the AHELO sample design .....	6
Introduction .....	6
Sampling requirements and expectations.....	6
Defining the target populations.....	7
Students in Generic Skills, Economics and Engineering strands.....	7
Level of programs .....	7
International definition of the student target population.....	8
International definition of the faculty target population .....	8
Coverage, exemptions and exclusions .....	9
National definition of the target population .....	10
Establishing the national sampling plan .....	10
Introduction .....	10
Working within the international sampling plan.....	11
Institution-specific issues in sample design .....	15
Participation rates.....	15
References.....	16
Appendix: Major operational steps for preparing the AHELO samples.....	17
Option 1: One-stage designs: working from sampling frames of individuals.....	17
Step 1: Preparing a complete sampling frame of individuals.....	17
Step 2: Preparing a working version of the sampling frame .....	19
Step 3: Sending the working version of the sampling frame to the NPM .....	20
Step 4: Selecting samples from the working version of the sampling frame .....	20
Option 2: Multi-stage designs: working from sampling frames of groups .....	24
Step 1: Preparing the sampling frame for the first stage of sampling .....	24
Step 2: Selecting the first-stage sample.....	25
Step 3: Preparing the sampling frame for the second stage of sampling.....	26
Step 4: Selecting the second-stage sample .....	27
Step 5: Preparing the sampling frame for the third stage of sampling .....	28
Step 6: Selecting the third-stage sample.....	30
Annex: ISCED Broad groups and fields of education.....	32

## **Introduction**

### ***About this AHELO Sampling Manual***

1. A study like the OECD Assessment of Higher Education Learning Outcomes (AHELO) is a major undertaking. Reliability and validity of the survey findings can only be achieved at the cost of extensive field work, thorough systems testing, rigorous instrument design and validation, sound sampling and estimation methods, and appropriate reporting.
2. It is important to emphasise that scientific sampling has not been used as widely in research relating to international comparisons in higher or post-compulsory education as it has, for instance, in school-level studies. Any study of the scale of AHELO, however, must make use of such methods. Non-random convenience or opportunity sampling methods are not sufficient to ensure the generalisability of estimates, and hence the comparability and validity of results. In other words, the successful application of probabilistic sampling methods is essential for the feasibility of AHELO.
3. This manual describes the steps required to acquire a statistical sample of students and faculty (see paragraphs 27 and following for proper definitions) so that the goals of the AHELO Feasibility Study be met. This manual is primarily intended for the National Project Managers (NPMs) and Institutional Coordinators (ICs) who will be responsible for the implementation of AHELO in their respective countries.
4. While some activities can only be carried out effectively and efficiently by NPMs and their representatives, others are best carried out by the AHELO Consortium. With the expected diversity of institutional arrangements of students, faculty members, departments and programs, activities related to the sampling of students and faculty members are thought to be best left to the care of national representatives.
5. Very few details about the actual sample selection procedures are given in the body of this document. NPMs or their representatives are referred to procedures and software algorithms provided in the Appendix.
6. In this document, the AHELO Consortium may be referred to as ‘the AHELO Sampling Team’ or the ‘Sampling Team’.

### ***AHELO’s goals***

7. As stated in one of the AHELO defining documents: “the purpose of the AHELO Feasibility Study is to assess whether it is possible to measure what undergraduate degree students know and can do at the international level, in order to provide relevant information to higher education institutions (HEIs), governments, and other stakeholders including students and employers” (OECD (2009a:2).
8. Since “the focus [of the AHELO feasibility study] is on providing proof of concept” (OECD (2009b: 15), and for several other reasons, it has been agreed that there will be no attempt at providing participating countries with nationally representative (probabilistic) samples of HEIs. This is an area that would require further investigation should there be a requirement to select HEIs using probabilistic sampling.

9. Attaining the goals laid out in paragraphs 7 and 8 entail conducting a sample survey of students in the final year of their respective undergraduate program and of faculty members of selected departments in the participating HEIs. For the purposes of the AHELO Feasibility Study, the term ‘program’ is defined as a specific pathway that exists within an institution that requires students to undertake a set of subjects and experiences, and leads to the award of a common credential or credentials on completion. The word ‘major’ is a common alternative term. Given that the participating HEIs will be recruited on a voluntary basis (OECD, 2009c: 9) for the Feasibility Study, recruitment of HEIs is not covered in this manual.
10. Besides the obvious requirements of being relevant to policy making and to establish the feasibility of conducting a full-scale AHELO, some quality standards are imposed on the strategy: validity, reliability, and rigour of the information produced, efficiency and cost-effectiveness of the assessment process.
11. A probability sampling design has been developed to reach the study goals while adhering to the quality standards listed above. Strict conformity to the principles and directions given in this manual is the key to achieving all those goals. Any deviation from the survey principles and directions must be fully justified, documented, and agreed to by the Sampling Team.

### ***Sampling plan for AHELO***

12. A probabilistic sample of students and faculty members will be selected within each participating HEI. The sampling plan for AHELO entails the following key steps:
  - a. Population of students:
    - i. Identifying the target population of students for which results are expected in each participating institution;
    - ii. Creating a list of students that covers the target population;
    - iii. Selecting a probability sample of students from that list; and
    - iv. Computing design-based estimation weights to allow unbiased HEI-level estimates of student characteristics and scores.
  - b. Population of faculty:
    - i. Identifying the target population of faculty for which results are expected in each of the participating institution;
    - ii. Creating a list of faculty members that covers the target population;
    - iii. Selecting a probability sample of faculty members from that list; and

- iv. Computing design-based estimation weights to allow unbiased HEI-level estimates of faculty characteristics and scores.
13. Progressing through these key steps will be the result of a fruitful collaboration between NPMs and the AHELO Consortium. The respective roles and responsibilities of each partner are described below.

***NPM responsibilities in sampling***

14. AHELO's NPMs are responsible for these tasks:
- a. Obtaining or creating the list of students and of faculty members in each participating HEI with support from the Institution Coordinator (IC);
  - b. Selecting the sample of students and faculty members in each participating HEI according to agreed upon procedures;
  - c. Keeping track of students' and faculty members' participation, and documenting non-participation with support from the IC; and
  - d. Contacting the Sampling Team to resolve any issue regarding sampling and ask for assistance in the design and implementation of the assessment.

***Sampling Team responsibilities***

15. The AHELO Sampling Team is responsible for the following tasks:
- a. Developing an international sampling plan and helping adapt it to the conditions of participating countries and HEIs;
  - b. Developing sampling algorithms;
  - c. Developing estimation and replication weights for each sample in each HEI;
  - d. Developing and measuring the appropriate exclusion, coverage, response and participation rates, for students and faculty members, as required; and
  - e. Assisting in the evaluation of the quality of data from each HEI.
16. While the sampling plan prepared for AHELO is quite general and flexible, the Sampling Team fully realises that some participating HEIs may find it difficult or impossible to abide by all the requirements set in this Sampling Manual. NPMs are invited to contact the Sampling Team to discuss the particular conditions under which they will have to work (e.g. programs offered by the institutions, sample co-ordination) and the acceptable adaptation of the general sampling plan to their HEIs' conditions.

## **Overview of the AHELO sample design**

### ***Introduction***

17. The AHELO Feasibility Study comprises three assessment strands: Generic Skills, Economics and Engineering. For each strand, a random sample of faculty and students will be invited to participate in AHELO.
18. Each participating country will select a convenience sample of (typically) 10 institutions for each strand in which they have opted to participate. These institutions should reflect the diversity of the higher education system and, for countries taking part in the discipline strands, offer first-cycle undergraduate programs in economics or engineering.
19. Probability samples will be selected within each institution. This means that individual participants are to be randomly selected from the list of in-scope potential participants for each population of interest in the selected institutions. Stratification may be used to ensure that certain sub-populations of interest are sufficiently represented to allow estimates of adequate precision.
20. In each institution, the faculty are randomly selected for the Generic Skills strand, but given the potentially small population size they are likely to all be selected for the Economics and Engineering strands.
21. The target populations of interest for the AHELO Feasibility Study are described in the next section.

### ***Sampling requirements and expectations***

22. Survey estimates will be required for students and faculty within participating institutions. To obtain reliable estimates for each of these analytical units, quality standards must be maintained with respect to response rates, sample size and national implementation:
  - a. Response rates: To maintain a minimum level of quality and minimize the effect of non-response biases, at least 75% of students within an institution's sample are expected to participate in the survey.
  - b. Sample size for the sample of students:
    - i. The target sample size – the students invited to participate in the assessment – is 200 students per participating institution and strand, regardless of the strand, to account for an expected loss of 25% to non-response. The expected yield is 150 participating students.
    - ii. Sample size may be adapted to actual enrolment and institutional contexts. For example, it may be judicious to survey all if the enrolment is 225 or less.
  - c. Sample size for the sample of faculty:
    - i. The target sample size is 40 faculty members per participating institution and strand, regardless of the strand, to account for an



expected loss of 25% to non-response. Hence the expected yield is 30 participating faculty. These numbers are based on statistical power and effect size considerations.

- ii. Sample size may be adapted to the actual number of eligible faculty. For example, it may be judicious to invite all faculty to participate if there are 50 or less.
  
- d. National implementation: In some HEIs, the number of students or number of faculty may be only slightly larger than the sample size requirements for AHELO Feasibility Study. In such cases, the NPMs should discuss with the Sampling Team to devise an adaptation of the international sampling plan suitable to the country or HEI. Countries participating in the Generic Skills strand may wish to optimise the sample of students drawn from the participating HEIs. Such options should be discussed with the Sampling Team.
  
- e. To help maintain the statistical properties of the random samples and to minimise the risks of ‘volunteer sampling’, replacement of non-respondents (students or faculty) is not allowed.

## **Defining the target populations**

### ***Students in Generic Skills, Economics and Engineering strands***

23. The AHELO Feasibility Study will survey students who are at the end of a three- or four-year undergraduate degree in an Economics program (Economics strand), an Engineering program (Engineering strand) or in any three- or four-year undergraduate program offered by the participating HEIs (Generic Skills strand).
  
24. For operational purposes the term ‘student’ refers to a person who is enrolled, follows a course of study or attends classes at a HEI.

### ***Level of programs***

25. The NPM has to inform the AHELO Sampling Team how the levels of programs are defined in his/her country. The following statements made by the Group of National Experts (GNE) in 2009 provides guidance in outlining the program levels of interest (OECD, 2009):

*“50. [...]The programs are referenced by the UNESCO International Standard Classification of Education (ISCED) as level 5A or 5B.[...] Programs can also be referenced at level 6 of the European Qualification Framework [...].*

*Although the ISCED level 5 programs’ duration typically ranges from less than 3 years to up to 6 years, the appropriate focus for the feasibility study is on the first range category, i.e. programs with a total cumulative duration of 3 to 4 years. Depending on countries, the first degree may last 3 years (e.g. Bologna-type degree structure based on 180 ECTS [European Credit Transfer and Accumulation System]) or 4 years in Mexico, Japan, Korea and Australia. These first programs are typically named Bachelor in English-speaking countries, Honours in Australia, Gakushi in Japan, Licentura in Mexico, Haksa in Korea). European countries have opted for a wording equivalent to Bachelor (Laurea in Italy for instance). [...]*

54. *The treatment of multidisciplinary programs is slightly problematic. [...] In deciding whether to consider the program for participation in the AHELO feasibility study, the amount of credits (in terms of duration for instance) of the selected disciplines should prevail. As a rule, students should significantly major in the selected discipline (e.g. 2/3 of the degree FTE credits should be spent studying the major subject) in order to qualify for inclusion in the target population for the AHELO discipline-specific assessments.*"

### ***International definition of the student target population***

26. The AHELO Feasibility Study intends to cover all students who are at the end of a three- or four-year undergraduate degree in a participating HEI. Students are considered in their final year of study if they are expected to graduate or obtain their qualification if they successfully complete all the classes they are enrolled in. Exclusions and exemptions are given below.
27. The target population for the Generic Skills strand comprises all students at the end of a three- or four-year undergraduate degree in one of the programs given by their institution.
28. The target population for the Economics strand comprises all students at the end of a three- or four-year undergraduate degree offered by the Economics department or faculty. It also comprises all students at the end of a three- or four-year undergraduate degree in a multidisciplinary program, who significantly majored in Economics (OECD, 2009: §54; see 25 above).
29. The target population for the Engineering strand comprises all full-time students at the end of a three- or four-year undergraduate degree in civil engineering of the Engineering department or faculty. It also comprises all full-time students at the end of a three- or four-year undergraduate degree in a multidisciplinary program, who significantly majored in Civil Engineering (OECD, 2009: §54; see 25 above).

### ***International definition of the faculty target population***

30. AHELO intends to cover all faculty members whose primary appointment is in an academic or interdisciplinary unit and whose responsibilities include teaching undergraduates. The term 'academic staff', 'teaching staff' or 'faculty' covers a wide range of different roles. An incomplete list includes: Residential Tutor, Assistant Lecturer, Lecturer, Senior Lecturer, Associate Professor, Professor, Clinical Supervisor, Research Assistant, Research Fellow, Senior Research Fellow, Honorary Fellow, Sessional Lecturer, Sessional Tutor, Marker, Examiner, Supervisor, Reader, Principal Research Fellow, Professorial Fellow, Postdoctoral Fellow, Head, Chair, Dean, Director, Deputy Vice-Chancellor, Provost, Pro Vice-Chancellor, Vice-Chancellor and Chancellor. Exclusions and exemptions are given below.
31. The target population for the Generic Skills strand comprises all teachers, lecturers and professors responsible for one or many undergraduate classes or sections offered to the students enrolled in one of the programs of their respective departments during the academic year covering the collection period (for example, a philosophy professor teaching logic to philosophy undergraduates is included, but the philosophy professor who only teaches courses outside the philosophy department, e.g. to science undergraduates, would be considered out-of-scope).

32. The target population for Economics strand comprises all teachers, lecturers and professors members of the Economics department / faculty responsible for one or many undergraduate classes or sections offered to the students enrolled in the in-scope Economics programs of the Economics department / faculty (e.g. the economics professor teaching micro-economics to Economics students is included but not the history professor teaching history of economics taken by the same undergraduates, nor the economics professors teaching courses only to students outside of the Economics department / faculty) during the academic year covering the collection period.
33. The target population for the Engineering strand comprises all teachers, lecturers and professors members of the Engineering department / faculty responsible for one or many undergraduate classes or sections offered to the students enrolled in Civil Engineering of the Engineering department / faculty (e.g. the engineering professor teaching engineering to students in civil engineering is included but not the management professor teaching a “project accounting” class taken by the same undergraduates, nor the Engineering professors teaching courses only to students outside of Civil Engineering) during the academic year covering the collection period.

#### *Coverage, exemptions and exclusions*

34. Clearly, this feasibility study must attempt to cover as much of the target populations as is politically, conceptually, economically and practically possible. Because of their unique workload, responsibilities or means of access to education, certain groups of individuals (faculty and students alike) are to be **excluded**, that is to say that **they are not to be counted as part of the target population, nor are they to be made eligible for sampling**. Estimates for their institution would not apply to them nor cover them.
35. Because of their unique characteristics, because they would require special arrangements that may be too expensive or difficult to organise for a feasibility study, or because of their duties, certain groups or persons may be **exempted from collection**. However, they are considered in-scope. Contrary to excluded persons or groups, these persons are to be listed and subjected to sampling and **if selected, may be exempted from having to sit through a session**. Estimates for their institution would pertain to and cover them.
36. Students in a cooperative profile who are in internship during the collection period, students in an international profile who are in a foreign country during the collection period, students studying via an external/distance mode are **excluded**. Students with special needs (e.g. visual impairment, reduced mobility) may be **exempted** if selected.
37. Following the GNE, “programs directed at adult publics which are equivalent in content with more traditional ISCED 5 programs should be covered” (OECD, 2009: §53).
38. Part-time teachers, lecturers or professors, replacement or invited teachers, lecturers or professors, research academics and senior executives without current teaching roles, honorary staff, and administrative staff are **excluded**.
39. For all strands, teachers, lecturers or professors who teach courses only to students outside of their respective department are **excluded**.

40. For all strands, teachers, lecturers or professors who teach courses within their respective departments but who have to assume a significant amount of administrative work (e.g. reduced teaching load during their tenure as Department Head) may be **exempted** if selected to avoid undue response burden.

### *National definition of the target population*

41. The national (institutional) definition of the target population must meet the requirements of the international target population. The national definition of the target population should provide the most exhaustive institutional coverage of students and faculty as is feasible. However, political, organizational or operational reasons might make it extremely difficult for some NPMs to obtain complete institutional coverage. Careful records should be made of identifiable exclusions and exemptions.
42. In accordance with the international target population, NPMs must define, describe and report on the target population for each participating HEI. If the institutional target population must deviate from the international standard, the deviations and ensuing loss of coverage must be documented. This situation will occur when a particular group of students and faculty must be removed from the target population for political, operational or administrative reasons. NPMs are encouraged to keep these exclusions and exemptions to a minimum. If the size of the excluded groups is significant, the survey results will not be representative of the complete target population and should be noted as such in the final reports and tables. The NPMs must list the reasons for exclusions, if any, as well as the size of the excluded population in terms of the number of students and faculty.
43. Within institutions, all in-scope students and faculty should be listed. Criteria for inclusion are given below. The students and faculty who are to be excluded from the sample for one of the reasons given above should be listed last and marked as exclusions.
44. NPMs should note that selected students and faculty absent for the duration of the assessment period but who would likely be present if the collection period were shifted a little (e.g. absent due to short-term illness) are not exclusions but are rather non-respondents. Selected students and faculty who would refuse or would otherwise be unable to take part in the assessment are also to be considered as non-respondents. However, in-scope individuals who are unavailable for the duration of the assessment regardless of the actual collection window (e.g. parental or sabbatical leave, internship) are **excluded**. Exclusions are to be determined before the sample of students and faculty is selected and the assessment takes place. Exclusions cannot be a by-product of the assessment.

## **Establishing the national sampling plan**

### *Introduction*

45. The Sampling Team will assist NPMs develop an efficient and effective sampling plan adapted to the conditions of their participating HEIs. Yet, the definition of the institutional conditions, the determination of the relevant stratification variables and the identification of possible impediments are the responsibility of each NPM. All institutional sampling plans will be developed jointly by the concerned NPM and the

Sampling Team and must be approved by the Sampling Team on behalf of the AHELO Consortium.

### *Working within the international sampling plan*

46. The basic international sampling plan developed for AHELO is a single-stage random sampling design. Any modification to the general design contemplated by a participating institution must be approved by the Sampling Team. The principal elements of the international plan, and some of the possible variations, are described in this section. When the basic design cannot be implemented, a multi-stage design, as described below, is an acceptable alternative.
47. All sampling activities rely on the existence of a complete, exhaustive and current sampling frame. The NPMs are responsible for the acquisition or the creation of such a sampling frame for the students and faculty in each participating HEI. The Sampling Team will provide further details on the characteristics of the sampling frame.
48. Within each participating institution, the IC within each HEI will prepare the following lists:
  - a. Generic Skills strand:
    - i. Sample of students: The IC will prepare a list of all in-scope students. An adequate sorting and sampling strategy would provide a better coverage of programs, gender and age. A systematic equal probability sample of 200 students will be drawn from the implicitly stratified list of in-scope students. Systematic sampling (Kish, 1965: 20-22; Lohr, 1999: 42-43) of students will provide a fair (approximately proportional to program enrolment) coverage of the implicit strata. Explicit stratification can be implemented where justified by the number of in-scope students and particular analytical requirements (see paragraph 19 above). Alternatively, under certain conditions, lists of classes or courses could be used as a proxy to student rosters. For reference, implicit stratification refers to sorting the file by some characteristic and then drawing a systematic random sample from the full file. This strategy should yield approximately proportional representation of the characteristic of interest. Explicit stratification means that each stratum is treated as an independent population and sample sizes can be fixed as required.
    - ii. Sample of faculty: The IC will prepare a list of all in-scope faculty. An adequate sorting and sampling strategy would provide a better coverage of programs, gender and age where justified by the number of in-scope faculty. A systematic equal probability sample of 40 faculty will be drawn from the implicitly stratified list of in-scope faculty. Systematic sampling of faculty will provide a fair coverage of the implicit strata. Explicit stratification can be implemented where justified by the number of in-scope faculty.



b. Economics strand:

- i. Sample of students: The IC will prepare a list of all in-scope students. An adequate implicit stratification would provide a better coverage of economics programs, for instance gender and age. A systematic equal probability sample of 200 students will be drawn from the implicitly stratified list of in-scope students. Systematic sampling of students will provide a fair (approximately proportional) coverage of the implicit strata. Explicit stratification can be implemented where justified by the number of in-scope students. Alternatively, under certain conditions (similar to those given in paragraph 50), lists of classes or courses could be used as a proxy to student rosters. Implicit and explicit stratification are defined above.
- ii. Sample of faculty: The IC will prepare a list of all in-scope faculty. An adequate implicit stratification would provide a better coverage of gender, age and academic specialty where justified by the number of in-scope faculty. A systematic equal probability sample of 40 faculty will be drawn from the implicitly stratified list of in-scope faculty. Systematic sampling of faculty will provide a fair coverage of the implicit strata. Explicit stratification can be implemented where justified by the number of in-scope faculty.

c. Engineering strand:

- i. Sample of students: The IC will prepare a list of all in-scope students. Implicit stratification may provide a better coverage of characteristics such as gender and age, for instance, of the students in civil engineering. A systematic equal probability sample of 200 students will be drawn from the implicitly stratified list of in-scope students. Systematic sampling of students will provide a fair (approximately proportional) coverage of the implicit strata. Explicit stratification can be implemented where justified by the number of in-scope students. Alternatively, under certain conditions (similar to those given in paragraph 50), lists of classes or courses could be used as a proxy to student rosters. Implicit and explicit stratification are defined above.
- ii. Sample of faculty: The IC will prepare a list of all in-scope faculty. An adequate implicit stratification would provide a better coverage of gender, age and academic specialty where justified by the number of in-scope faculty. A systematic equal probability sample of 40 faculty will be drawn from the implicitly stratified list of in-scope faculty. Systematic sampling of faculty will provide a fair coverage of the implicit strata. Explicit stratification can be implemented where justified by the number of in-scope faculty.

49. Lists of students and faculty in participating institutions are to be as current as possible at the time of the sampling in order to minimise non-response and coverage biases.

50. In some institutions, there will be no comprehensive list of students available to the IC, or available list of students may not provide the programs in which the students are enrolled. If time and resources do not allow the IC to construct a HEI-level comprehensive list (for the Generic Skills strand) or program-specific lists (for the Economics and Engineering strands), a different design could be implemented as explained below.

a. Generic Skills strand three-stage design:

- i. First stage – a sample of programs. The IC will prepare a list of all undergraduate programs offered by the participating institution. Ten programs will be randomly selected from that list. Programs may be selected with unequal probabilities if information on program enrolment is available.
- ii. Second stage – a sample of courses: For each of the selected programs, the IC will prepare the list of the compulsory courses given to in-scope students. Only courses that have to be taken by all in-scope students should appear on the list of in-scope courses. This ensures that all in-scope students are accounted for. Then, for each selected program, one course is randomly selected from the list of courses. Where such courses do not exist, a combination of compulsory classes can attain the same objective. This deviation has to be discussed by the concerned NPM and the Sampling Team.
- iii. Third stage – a sample of students:
  1. Option 1: The nominal sample size of 200 students will be divided by the number of selected programs, so that an equal number of students be selected from each participating program, likely 20 students per program. For each of the selected courses in the selected programs, the list of in-scope students must be compiled or obtained. The 10 random samples of 20 students can be drawn from those lists. Barring non-response, this strategy will yield a self-weighted sample.
  2. Option 2: The nominal sample size of 200 students will be divided proportionally among the selected programs, so that a proportional number of students will be selected from each participating program. This option will tend to favour the more popular programs. A minimum sample size per program could be required, for instance, 10. The IC will prepare a list of all in-scope students from the selected course(s). Random sample of students will be drawn from the lists of in-scope students.

b. Economics strand stratified single-stage design:

- i. Stratification: If desired, each Economics program could constitute a stratum. In each stratum or program, the IC will prepare the list of the compulsory courses given to in-scope students. If such stratification is not desirable, then the full cohort forms a single stratum. Only courses that have to be taken by all in-scope students should appear on the list of in-scope courses. This ensures that all in-scope students are accounted for. Then, for each program, one course is randomly selected from the list of courses. Since each course on that list is expected to comprise all in-scope students, the random draw on one such course is inconsequential. Where such courses do not exist, a combination of compulsory classes can attain the same objective. This deviation has to be discussed by the concerned NPM and the Sampling Team.
  - ii. Sampling students: (option 1) The nominal sample size of 200 students will be divided by the number of programs, so that an equal number of students be selected from each program. Barring non-response, this strategy will yield about the same accuracy for each program, assuming that the variability of the characteristics of interest is about the same across programs. (option 2) The nominal sample size of 200 students will be divided proportionally among the programs. This option will tend to favour the more popular program. This strategy will not yield nearly equal accuracy for program-specific statistics but will give better accuracy overall than option 1. The IC will prepare a list of all in-scope students from the selected course(s). Implicit stratification may provide a better coverage of characteristics such as gender and age. A systematic equal probability sample of  $n$  ( $n$  depends on program enrolment) students will be drawn from the implicitly stratified list of in-scope students. Alternatively, under certain conditions, entire classes or courses could be selected in lieu of sampling individual students.
  - iii. In the selected courses, students who are not in-scope are to be reported on the list but marked as out-of-scope (e.g. a student in the physics program who takes a course in the economics program, or a student who would not graduate at the end of the year).
- c. Engineering strand stratified single-stage design:
- i. Stratification: The Civil Engineering program constitutes its own stratum. For that program, or stratum, the IC will prepare the list of the compulsory courses given to in-scope students. Only courses that have to be taken by all in-scope students should appear on the list of in-scope courses. This ensures that all in-scope students are accounted for. Then, one course is randomly selected from the list of courses. Where such courses do not exist, a combination of compulsory classes can attain the same objective. This deviation has to be discussed by the concerned NPM and the Sampling Team.



- ii. Sampling students: The IC will prepare a list of all in-scope students from the selected course(s). Implicit stratification may provide a better coverage of characteristics such as gender and age. A systematic equal probability sample of  $n$  ( $n$  depends on the program enrolment) students will be drawn from the implicitly stratified list of in-scope students. Alternatively, under certain conditions, entire classes or courses could be selected in lieu of sampling individual students.
- iii. In the selected courses, students who are not in-scope are to be reported on the list but marked as out-of-scope (e.g. a student in the physics program who takes a course in the civil engineering program, a mechanical engineering student taking the selected compulsory class of civil engineering as an elective, or a student who would not graduate at the end of the year).

### ***Institution-specific issues in sample design***

- 51. The following paragraphs may not apply to every participating institution. They address particularities shared by few participants. NPMs that are faced by one of these issues or some other unique situation are invited to contact the Sampling Team to discuss a resolution suitable to their context.
- 52. The sample sizes given above have been calculated for institutions with what may be considered to be a moderate or large number of students and faculty.
- 53. The students sample size is set at 200 to allow for some amount of non-response. In institutions (Generic Skills strand) or programs (Economics and Engineering strands) where fewer than 200 in-scope students can be found, all students should be selected.
- 54. The faculty sample size is set at 40 to allow for some amount of non-response. In institutions (Generic Skills strand) or programs (Economics and Engineering strands) where fewer than 40 in-scope faculty can be found, they should all be selected.
- 55. Within the sampling framework prepared for AHELO, many national options are possible. NPMs must have their options approved by Statistics Canada on behalf of the AHELO Sampling Team to ensure that:
  - a. they do not interfere with the goals, objectives and expectations of AHELO;
  - b. they are statistically sound; and
  - c. their sampling plans are adapted to both the international requirements and the national options.

### ***Participation rates***

- 56. The quality standard requirements for AHELO sampling are given earlier in this document (see paragraph 22). These requirements translate in participation rates (response rates) for students and faculty members. Reaching these levels of participation does not preclude that some amount of error may be present in the results but should

minimise the negative impact of non-response biases. Recent international studies conducted on undergraduates (Dumais, Meink, in press) would indicate that participation of university faculty ranges from 60% to 98% whereas the participation of last-year undergraduates ranges from 75% to 99% depending on the country (countries where participation was below 50% are not taken into account).

57. AHELO expects that at least 75% of the selected students and faculty members in participating institutions take part in the assessment.
58. Both weighted and unweighted participation rates for students and faculty will be produced for each HEI. More information will be provided about the concept and practice of ‘weighting’ in other documents and workshops.

## References

Dumais, J., & Meinck, S. (in press). “Weighting and Variance Estimation”. in Tatto, M.T., Schulle, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G., editors, *Teacher Education and Development Study in Mathematics (TEDS-M): Technical Report*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University (to appear).

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

Lohr, S. (1999). *Sampling: Design and analysis*. Pacific Grove: Duxbury Press.

Organisation for Economic Cooperation and Development (OECD) (2006). *Sampling Manual, Survey of Teachers, Teaching and Learning (TALIS), Field Test Version*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD) (2009). *Feasibility Study for an Assessment of Higher Education Learning Outcomes (AHELO)*. OECD: Paris.

Organisation for Economic Cooperation and Development (OECD) (2009a). *Roadmap For The OECD Assessment Of Higher Education Learning Outcomes (AHELO) Feasibility Study*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD) (2009b). *AHELO Terms of Reference*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD) (2009c). *Role of National Project Managers and resource needs*. OECD: Paris.

Organisation for Economic Cooperation and Development (OECD) (2010). *Assessment Design: Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study*. Paris: OECD.

United Nations Educational, Scientific and Cultural Organisation (UNESCO) (2006). *ISCED 1997: International Standard Classification of Education ISCED 1997*. Montreal: unesco Institute for Statistics.

## Appendix: Major operational steps for preparing the AHELO samples

### *Option 1: One-stage designs: working from sampling frames of individuals*

The following paragraphs describe the key steps to obtain a probabilistic sample of individuals from a population list of those individuals, that is, a sampling frame. The list must be as complete and as current as technically possible in order to minimise coverage biases. These steps apply equally to the sampling of students and to the sampling of faculty.

Coordination with institutional or departmental registry staff is a key element of obtaining the required lists. When a single, unduplicated list of individuals cannot be obtained from a central (institutional level) registry, a combination of partial lists may achieve the same purpose. The IC will need to ensure that the partial lists do cover the entire in-scope population of interest. The IC will also need to ensure that the partial lists are unduplicated and combined into a single frame of in-scope individuals.

#### *Step 1: Preparing a complete sampling frame of individuals*

The IC will obtain or assemble one sampling frame for each strand in which her or his institution is invited to participate by:

- Ensuring each in-scope individual appears only once on the sampling frame. The IC will run a check to ensure that duplicate records are removed;
- Adding data for these variables: GENDER (student gender); STUDAPT (a measure of student aptitude/achievement available to the institution); BROADFIELD (broad field of study defined using ISCED, to assist with sampling and test administration – see the Annex); NARROWFIELD (narrow field of study defined using ISCED, to assist with sampling and test administration – see the Annex); NAME (student name to assist with test administration); INSTUDID (the institution's student ID, to assist with test administration); DOB (student date of birth, for sampling and test administration); EMAIL (student electronic mail address, to assist with test administration); and PHONE (student phone number, to assist with test administration);
- Adding up to three **implicit** strata to the frame (IMPSTR1 to IMPSTR3) to the frame for sorting prior to sample selection if they are thought to improve the mixture of respondents in the sample or to be useful as calibration information during estimation (for instance department, program, ethnicity, mother tongue, residency or immigration status);
- Adding one explicit stratum identifier (EXPLSTR) and the sample size allocated to each stratum (STRSAMP) if **explicit** stratification is to be used. Strata can be formed using a single variable (e.g. department) or by crossing several variables (e.g. department, gender, range of high school GPA) – see paragraph 48. The stratum identifier and the sample size associated to the stratum **must** be supplied for each individual on the sampling frame to allow for explicit stratification. Specifications for the definition of explicit strata should come from the NPM, likely as the result of discussions between the IC and the NPM;

- Adding a STATUS for each individual in the frame to indicate who is **excluded** from the population or **exempted** from collection (see paragraphs 34 to 40); and
- Saving the complete sampling frame and keeping it safe for use at a later stage. The file name should clearly indicate for which institution and which population that the file is the complete sampling frame.
- The suggested names and formats are :
  - i. <INSTID>STUDENT\_CFRAME.xls or \*.xlsx if saved under Microsoft Excel; or
  - ii. <INSTID>STUDENT\_CFRAME.sav for SPSS; or
  - iii. I<INSTID>STUDENT\_CFRAME.sas7bdat for SAS as SAS will not accept file names starting with a number; or
  - iv. <INSTID>FACULTY\_CFRAME.xls or \*.xlsx (Microsoft Excel); or
  - v. <INSTID>FACULTY\_CFRAME.sav (SPSS); or
  - vi. I<INSTID>FACULTY\_CFRAME.sas7bdat (SAS), where <INSTID> is the institution identification number assigned to a participating institution be the Consortium; for example the file I0142FACULTY\_CFRAME is the SAS version of the complete frame of faculty members for institution number 0142.

At minimum, the complete version of the sampling frame should comprise the variables shown in Table 1, one record per individual.

**Table 1: Variables in the complete version of the student-level sampling frame**

Variable	Source	Label	Value	Format	Other
FILEID	Consortium	Data file ID	1 to n	F10.0	
COUNTRYID	Consortium	Country ID	ISO Country Code	F3.0	
REGIONID	Consortium	Sub-region ID	ISO Country Code	F5.0	
STRANDID	Consortium	Strand ID	1=Generic Skills 2=Economics 3=Engineering	F2.0	
INSTID	Consortium	Institution ID	0001 to 9999	F4.0	
STUDID	Consortium	Student ID	00001 to 99999	F5.0	
PASSWORD	Consortium	Test system password	Alphanumeric	A6	
GENDER	NPM, IC	Student gender	1=Male 2=Female	F1.0	7=N/A 8=Invalid 9=Missing
STUDAPT	NPM, IC	Student aptitude measure	Open text response	A100	
BROADFIELD	NPM, IC	Broad field of study	ISCED code (see Annex)	F1.0	
NARROWFIELD	NPM, IC	Narrow field of study	ISCED code (see Annex)	F2.0	
NAME	NPM, IC	Student name	Open text response	A100	7=N/A 8=Invalid 9=Missing
INSTUDID	NPM, IC	Institution student ID	Open text response	A100	7=N/A 8=Invalid 9=Missing
DOB	NPM, IC	Student date of birth	YYYYMMDD	F8.0	7=N/A 8=Invalid 9=Missing
EMAIL	NPM, IC	Student electronic mail address	Open text response	A50	7=N/A 8=Invalid 9=Missing

PHONE	NPM, IC	Student phone number	Open text response	A50	7=N/A 8=Invalid 9=Missing
STATUS	NPM, IC	Student eligibility status	1=Included 2=Exempted 3=Excluded	F1.0	7=N/A 8=Invalid 9=Missing
IMPSTR1	NPM, IC	Implicit strata 1 identifier	Open text response	A100	
IMPSTR2	NPM, IC	Implicit strata 2 identifier	Open text response	A100	
IMPSTR3	NPM, IC	Implicit strata 3 identifier	Open text response	A100	
EXPLSTR	NPM, IC	Explicit strata identifier	Open text response	F2.0	
STRSAMP	NPM, IC	Explicit strata sample size	Open text response	F5.0	

### Step 2: Preparing a working version of the sampling frame

The IC will prepare a lighter, simpler, **anonymous** working version of the sampling frame, by:

- Removing all duplicate entries if any are left;
- Removing **excluded** individuals;
- Removing direct identifiers (e.g. NAME and INSTUDID) and contact information (e.g. EMAIL and PHONE); and
- Saving the working version of the sampling frame and keeping it safe for use at a later stage. The file name should clearly indicate that the file is the working version of the sampling frame.
- The suggested NAMES and FORMATS are
  - <INSTID>STUDENT\_WFRAME.xls or \*.xlsx (Microsoft Excel); or
  - <INSTID>STUDENT\_WFRAME.sav for SPSS files; or
  - I<INSTID>STUDENT\_WFRAME.sas7bdat for SAS files; and
  - <INSTID>FACULTY\_WFRAME.xls or \*.xlsx (Microsoft Excel); or
  - <INSTID>STUDENT\_WFRAME.sav (SPSS); or
  - I<INSTID>STUDENT\_WFRAME.sas7bdat (SAS).

The working version of the sampling frame should comprise the variables listed in Table 2, one record per individual.

**Table 2: Variables in reduced working version of student-level sampling frame**

Variable	Source	Label	Value	Format	Other
FILEID	Consortium	Data file ID	1 to n	F10.0	
COUNTRYID	Consortium	Country ID	ISO Country Code	F3.0	
REGIONID	Consortium	Sub-region ID	ISO Country Code	F5.0	
STRANDID	Consortium	Strand ID	1=Generic Skills 2=Economics 3=Engineering	F2.0	
INSTID	Consortium	Institution ID	0001 to 9999	F4.0	
STUDID	Consortium	Student ID	00001 to 99999	F5.0	
PASSWORD	Consortium	Test system password	Alphanumeric	A6	
GENDER	NPM, IC	Student gender	1=Male 2=Female	F1.0	7=N/A 8=Invalid 9=Missing
STUDAPT	NPM, IC	Student aptitude measure	Open text response	A100	
BROADFIELD	NPM, IC	Broad field of study	ISCED code (see Annex)	F1.0	

NARROWFIELD	NPM, IC	Narrow field of study	ISCED code (see Annex)	F2.0	
DOB	NPM, IC	Student date of birth	YYYYMMDD	F8.0	7=N/A 8=Invalid 9=Missing
STATUS	NPM, IC	Student eligibility status	1=Included 2=Exempted 3=Excluded	F1.0	7=N/A 8=Invalid 9=Missing
IMPSTR1	NPM, IC	Implicit strata 1 identifier	Open text response	A100	
IMPSTR2	NPM, IC	Implicit strata 2 identifier	Open text response	A100	
IMPSTR3	NPM, IC	Implicit strata 3 identifier	Open text response	A100	
EXPLSTR	NPM, IC	Explicit strata identifier	Open text response	F2.0	
STRSAMP	NPM, IC	Explicit strata sample size	Open text response	F5.0	

*Step 3: Sending the working version of the sampling frame to the NPM*

To send the working version of the sampling frame to the NPM the IC should:

- Verify with their NPM that compressed or zipped files will be allowed through the NPM's email firewall;
- Protect the working version of the sampling frame with a password. The easiest way to do this is to compress/zip the file and supply a password during the compression process. The IC will use **AHELO2012!** as the password; and
- Send the compressed working version of the sampling frame to the NPM by the date specified by the NPM.

*Step 4: Selecting samples from the working version of the sampling frame*

The NPM will draw the required random samples by:

- Decompressing and storing the working version of the institutional sampling frame;
- Sorting the frame by any explicit and implicit stratification variables;
- Drawing the sample using one of the scripts given below (or an equivalent, if neither SPSS nor SAS are used – NB: Microsoft Excel “Data Analysis - Sampling” macro is **not** recommended for this application);
- Recording and saving the log of the sample selection runs (one log for the student sample and one log for the faculty sample for each strand and each institution, clearly named);
- For quality assurance purposes, uploading a copy of working version of the frame, the log (including the random seed used) and the resulting sample to your country's secure space on the AHELO Exchange; and
- Storing copies of the logs and samples for later reference and sending copies (compressed and password-protected) of the selected samples to the IC.

Figure 1 to Figure 4 present sample pieces of syntax that illustrates how to use SPSS and SAS to draw the AHELO samples. Various scenarios are explored. Figure 1 presents SPSS syntax for drawing a simple random sample of  $n=200$  students. The NPM will provide parameters highlighted in red. For documentation and quality assurance, the NPM will record the random seed used for each selection (for illustration purposes, the seed “**123454321**” was used here). The optional SEED=RANDOM (SPSS ComplexSamples), or equivalently omitting

the `SEED` statement (SAS/STAT), does not allow the reproduction of the sample selection; undocumented random seed values are not allowed.



```

GET FILE= 'DRIVE:\FOLDER\SIMPLE_WOR
             ESTIMATION=DEFAULT
  /SIZE VALUE=200
  /STAGEVARS INCLPROB(InclusionProbability)
             CUMWEIGHT(CumulativeWeight)
             POPSIZE(PopulationSize)
             SAMPSIZE(SampleSize)
             WEIGHT(SampleWeight) .

CSSELECT
  /PLAN FILE= 'DRIVE:\FOLDER\123454321
  /CLASSMISSING EXCLUDE
  /DATA RENAMEVARS
  /SAMPLEFILE OUTFILE= 'DRIVE:\FOLDER\

```

**Figure 1: SPSS syntax for the selection of a simple random sample of size  $n=200$**

To draw a systematic random sample of  $n=200$  students, it is sufficient to replace the statement `/METHOD TYPE=SIMPLE_WOR` with the statement `/METHOD TYPE=SYSTEMATIC`. To draw the sample of faculty, the appropriate file names and sample size must be inserted where required.

Figure 2 shows the SAS syntax to draw simple random samples of students.

```

LIBNAME DRIVE "DRIVE:\FOLDER";

PROC SORT
  DATA= DRIVE.I<INSTID>STUDENT_WFRAME
  OUT= I<INSTID>STUDENT_SORTED;
  BY IMPSTR1 IMPSTR2 IMPSTR3 BROADFIELD DOB STUDAPT GENDER;
RUN;

PROC SURVEYSELECT
  DATA= I<INSTID>STUDENT_SORTED
  METHOD=SRS
  N= 200
  SEED= 123454321
  OUT= DRIVE.I<INSTID>STUDENT_SAMPLE;
RUN;

```

**Figure 2: SAS syntax for the selection of a simple random sample of size  $n=200$**



To draw a systematic random sample of  $n=200$  students, it is sufficient to replace the statement `METHOD=SRS` with the statement `METHOD=SYS`. To draw the sample of faculty, the appropriate file names and sample size must be inserted where required.

The next figures illustrate selection of stratified random samples using the SPSS ComplexSamples module (Figure 3) or SAS Surveysselect procedure found in SAS/STAT (Figure 4).

```

GET FILE= 'DRIVE:\FOLDER\<INSTID>STUDENT_WFRAME.SAV' .

SORT CASES BY EXPLSTR (A) IMPSTR1 (A) IMPSTR2 (A)
              IMPSTR3 (A) BROADFIELD (A) DOB (A) STUDAPT (A) GENDER (A) .

CSPLAN SAMPLE
/PLAN FILE= 'DRIVE:\FOLDER\<INSTID>STUDENT.csplan'
/PLANVARS SAMPLEWEIGHT=SampleWeight_Final_
/PRINT PLAN
/DESIGN STRATA= EXPLSTR
/METHOD TYPE=SIMPLE_WOR
          ESTIMATION=DEFAULT
/SIZE VARIABLE=STR_SAMP
/STAGEVARS INCLPROB(InclusionProbability)
           CUMWEIGHT(CumulativeWeight)
           POPSIZE(PopulationSize)
           SAMPSIZE(SampleSize) .

CSSELECT
/PLAN FILE= 'DRIVE:\FOLDER\<INSTID>STUDENT.csplan'
/CRITERIA STAGES=1
          SEED=123454321
/CLASSMISSING EXCLUDE
/DATA RENAMEVARS
/SAMPLEFILE OUTFILE=DRIVE:\FOLDER\<INSTID>STUDENT_SAMPLE.sav'
/PRINT SELECTION.

```

**Figure 3: SPSS syntax for the selection of a stratified simple random sample**

To draw a stratified systematic random sample, it is sufficient to replace the statement `/METHOD TYPE=SIMPLE_WOR` with the statement `/METHOD TYPE=SYSTEMATIC`. To draw the sample of faculty, the appropriate file names and sample size must be inserted where required.

```

LIBNAME DRIVE "DRIVE:\FOLDER";

PROC SORT
  DATA= DRIVE. I<INSTID>STUDENT_WFRAME
  OUT= I<INSTID>STUDENT_SORTED;
  BY EXPLSTR IMPSTR1 IMPSTR2 IMPSTR3 BROADFIELD DOB STUDAPT GENDER;
RUN;

PROC SURVEYSELECT
  DATA= I<INSTID>STUDENT_SORTED
  STRATA= EXPLSTR
  METHOD= SRS
  N= STR_SAMP
  SEED= 123454321
  OUT= DRIVE. I<INSTID>STUDENT_SAMPLE;
RUN;

```

**Figure 4: SAS syntax for the selection of a stratified simple random sample**

To draw a stratified systematic random sample, it is sufficient to replace the statement `METHOD=SRS` with the statement `METHOD=SYS`. To draw the sample of faculty, the appropriate file names and sample size must be inserted where required.

### *Option 2: Multi-stage designs: working from sampling frames of groups*

In the rare situation in which a sampling frame of individuals is impossible to obtain or assemble, methods for sampling individuals are nonetheless required. The next few paragraphs describe how frames and samples can be put together for multi-stage sampling. The syntax presented for option 1 is modified below.

Close coordination with institutional or departmental registry staff is a key element of obtaining the required lists. It is assumed that neither a single and unduplicated list of individuals nor partial lists of individuals can be obtained from the custodians. For the sample of students, the strategy will be to by-pass those lists by sampling departments and one of the compulsory classes that the students in their final year must take. The issue of creating a list can then be reduced to the list of those who attend the selected class.

The actual practical conditions in each participating institution are likely to be different and the steps given below might not apply directly as described here. However, the following steps are detailed enough to allow adaptation to most practical cases. Adaptations to actual situations should be discussed by the NPM and the Sampling Team before implementation.

#### *Step 1: Preparing the sampling frame for the first stage of sampling*

The IC will prepare the sampling frame for the first stage of sampling by:

- Obtaining or compiling the list of faculties or departments (UNIT1, the smallest available administrative group, e.g. Chemistry Department) offering undergraduate programs;
- Naming the list with a clear name (e.g. <INSTID>DEPT\_FRAME) and saving the list for later use; and
- Sending the list to the NPM for the selection of a random sample of 10 departments.

The file <INSTID>DEPT\_FRAME frame should contain the variables listed in Table 3 at minimum, with one record per UNIT1.

**Table 3: Variables in the first stage sampling frame**

Variable	Source	Label	Value	Format
FILEID	Consortium	Data file ID	1 to n	F10.0
COUNTRYID	Consortium	Country ID	ISO Country Code	F3.0
REGIONID	Consortium	Sub-regionID	ISO Country Code	F5.0
STRANDID	Consortium	Strand ID	1=Generic Skills 2=Economics 3=Engineering	F1.0
INSTID	Consortium	Institution ID	0001 to 9999	F4.0
UNIT1	NPM, IC	Smallest available administrative group	Open text response	A100

*Step 2: Selecting the first-stage sample*

The NPM will select a random sample of 10 departments from the list provided by the IC by:

- Adapting the script given in Figure 5 and drawing the random sample of 10 units;
- Recording and saving the log of the sample selection runs (one log for each strand and each institution, clearly named e.g. <INSTID>DEPT\_SAMPLE);
- For quality assurance purposes, uploading a copy of working version of the frame, the log (including the random seed used) and the resulting sample to your country's secure space on the AHELO Exchange; and
- Storing copies of the logs and samples for later reference and sending copies of the selected samples to the respective ICs.

```

GET FILE='DRIVE:\FOLDER\<INSTID>DEPT_FRAME.SAV'.

SORT CASES BY UNIT1 (A).

CSPLAN SAMPLE
  /PLAN FILE='DRIVE:\FOLDER\<INSTID>DEPT.csplan'
  /PLANVARS SAMPLEWEIGHT=SampleWeight_Final_1
  /PRINT PLAN
  /DESIGN STAGELABEL='first'
  /METHOD TYPE=SIMPLE_WOR
  ESTIMATION=DEFAULT
  /SIZE VALUE=10
  /STAGEVARS INCLPROB(InclusionProbability_1)
  CUMWEIGHT(CumulativeWeight)
  POPSIZE(PopulationSize_1)
  SAMPSIZE(SampleSize_1)
  WEIGHT(SampleWeight_1).

CSSELECT
  /PLAN FILE='DRIVE:\FOLDER \<INSTID>DEPT.csplan'
  /CRITERIA STAGES=1
  SEED=123454321
  /CLASSMISSING EXCLUDE
  /DATA RENAMEVARS
  /SAMPLEFILE OUTFILE='DRIVE:\FOLDER\<INSTID>DEPT_SAMPLE.sav'
  /PRINT SELECTION.

```

**Figure 5: SPSS syntax for the selection of a simple random sample of 10 UNIT1 (departments or similar administrative structure)**

*Step 3: Preparing the sampling frame for the second stage of sampling*

The IC will prepare the sampling frame for the second stage of sampling by:

- Obtaining or compiling the list of undergraduate compulsory units offered in the final year (UNIT2 – with unit possible also referred to as ‘courses’, ‘subjects’) to in-scope students for each sampled UNIT1 found on the <INSTID>DEPT\_SAMPLE file received from the NPM;
- Obtaining the student subgroups (UNIT3) for each UNIT2;
- Naming the list with a clear name (e.g. <INSTID>CRS\_FRAME) and saving the list for later use; and
- Sending the list of UNIT2 (courses) and UNIT3 (sections) to the NPM for the selection of a random sample of 2 UNIT3 groups in each of the two selected UNIT2 groups.

The file <INSTID>CRS\_FRAME should contain the variables listed in Table 4 at minimum, one record per UNIT3.

**Table 4: Variables in the second stage sampling frame**

Variable	Source	Label	Value	Format
FILEID	Consortium	Data file ID	1 to n	F10.0
COUNTRYID	Consortium	Country ID	See lookup tables	F3.0
REGIONID	Consortium	Sub-region ID	See lookup tables	F5.0
STRANDID	Consortium	Strand ID	1=Generic Skills 2=Economics 3=Engineering	F1.0
INSTID	Consortium	Institution ID	0001 to 9999	F4.0
UNIT1	NPM, IC	Smallest available administrative group	Open text response	A100
UNIT2	NPM, IC	Unit of study (course, subject)	Open text response	A100
UNIT3	NPM, IC	Unit subgroups (course section)	Open text response	A100

*Step 4: Selecting the second-stage sample*

The NPM will select a random sample of two course sections from the list provided by the IC by:

- Adapting the script given in Figure 6 and drawing the sample of two course sections or groups;
- Recording and saving the log of the sample selection runs (one log for each strand and each institution, clearly named e.g. <INSTID>DEPT\_SAMPLE);
- For quality assurance purposes, uploading a copy of working version of the frame, the log (including the random seed used) and the resulting sample to your country's secure space on the AHELO Exchange;
- Storing copies of the logs and samples for later reference and sending copies of the selected samples to the respective ICs.

```

GET FILE= 'DRIVE:\FOLDER\<INSTID>COURSE_FRAME.SAV'.

SORT CASES BY UNIT1 (A) UNIT2 (A) UNIT3 (A).

CSPLAN SAMPLE
  /PLAN FILE= 'DRIVE:\FOLDER\<INSTID>COURSE.csplan'
  /PLANVARS SAMPLEWEIGHT=SampleWeight_Final_2
  /PRINT PLAN
  /DESIGN STAGELABEL='second'
  STRATA=UNIT1
  /METHOD TYPE=SIMPLE_WOR
  ESTIMATION=DEFAULT
  /SIZE VALUE=2
  /STAGEVARS INCLPROB(InclusionProbability_2)
  CUMWEIGHT(CumulativeWeight)
  POPSIZE(PopulationSize_2)
  SAMPSIZE(SampleSize_2)
  WEIGHT(SampleWeight_2).

CSSELECT
  /PLAN FILE= 'DRIVE:\FOLDER\<INSTID>COURSE.csplan'
  /CRITERIA STAGES=1
  SEED=123454321
  /CLASSMISSING EXCLUDE
  /DATA RENAMEVARS
  /SAMPLEFILE OUTFILE= 'DRIVE:\FOLDER\<INSTID>COURSE_SAMPLE.sav'
  /PRINT SELECTION.

```

**Figure 6: SPSS syntax for the selection of a (stratified) simple random sample of two UNIT3 (course sections) per selected UNIT1 (department or similar)**

*Step 5: Preparing the sampling frame for the third stage of sampling*

The IC will prepare the sampling frame for the third stage of sampling by:

- Obtaining or compiling the list of in-scope students enrolled in the selected course sections or groups found on the <INSTID>COURSE\_SAMPLE file received from the NPM;
- Naming the list with a clear name (e.g. <INSTID>STD\_CFRAME) and saving the list for later use; and
- Sending the list of students to the NPM for the selection of a random sample of 10 students in each of the 2 selected course sections.

The third-stage student **complete** frame should contain the variables listed in Table 5 at minimum, with one record per individual:

**Table 5: Variables in the third stage sampling frame**

Variable	Source	Label	Value	Format	Other
FILEID	Consortium	Data file ID	1 to n	F10.0	
COUNTRYID	Consortium	Country ID	See lookup tables	F3.0	
REGIONID	Consortium	Sub-region ID	See lookup tables	F5.0	
STRANDID	Consortium	Strand ID	1=Generic Skills 2=Economics 3=Engineering	F1.0	
INSTID	Consortium	Institution ID	0001 to 9999	F4.0	
UNIT1	NPM, IC	Smallest available administrative group	Open text response	A100	
UNIT2	NPM, IC	Unit of study (course, subject or program)	Open text response	A100	
UNIT3	NPM, IC	Unit subgroups	Open text response	A100	
STUDID	Consortium	Student ID	00001 to 99999	F5.0	
PASSWORD	Consortium	Test system password	Alphanumeric	A6	
GENDER	NPM, IC	Student gender	1=Male 2=Female	F1.0	7=N/A 8=Invalid 9=Missing
STUDAPT	NPM, IC	Student aptitude measure	Open text response	A100	
BROADFIELD	NPM, IC	Broad field of study	ISCED code (see Annex)	F1.0	
NARROWFIELD	NPM, IC	Narrow field of study	ISCED code (see Annex)	F2.0	
NAME	NPM, IC	Student name	Open text response	A100	7=N/A 8=Invalid 9=Missing
INSTUDID	NPM, IC	Institution student ID	Open text response	A100	7=N/A 8=Invalid 9=Missing
DOB	NPM, IC	Student date of birth	YYYYMMDD	F8.0	7=N/A 8=Invalid 9=Missing
EMAIL	NPM, IC	Student electronic mail address	Open text response	A50	7=N/A 8=Invalid 9=Missing
PHONE	NPM, IC	Student phone number	Open text response	A50	7=N/A 8=Invalid 9=Missing
STATUS	NPM, IC	Student eligibility status	1=Included 2=Exempted 3=Excluded	F1.0	7=N/A 8=Invalid 9=Missing

To maintain anonymity, the IC will prepare a **working version** of the student frame by:

- Removing direct identifiers (e.g. NAME and INSTUDID) and contact information (e.g. EMAIL and PHONE); and
- Removing **excluded** individuals; and
- Saving the working version of the sampling frame and keeping it safe for use at a later stage; the file name should clearly indicate that the file is the working version of the sampling frame; the suggested name is <INSTID>STD\_WFRAME.

The working version of the sampling frame should comprise the variables listed in **Table 6**, with one record per individual.

**Table 6: Variables in the reduced working version of the third stage sampling frame**

Variable	Source	Label	Value	Format	Other
FILEID	Consortium	Data file ID	1 to n	F10.0	
COUNTRYID	Consortium	Country ID	See lookup tables	F3.0	
REGIONID	Consortium	Sub-region ID	See lookup tables	F5.0	
STRANDID	Consortium	Strand ID	1=Generic Skills 2=Economics 3=Engineering	F1.0	
INSTID	Consortium	Institution ID	0001 to 9999	F4.0	
UNIT1	NPM, IC	Smallest available administrative group	Open text response	A100	
UNIT2	NPM, IC	Unit of study (course, subject or program)	Open text response	A100	
UNIT3	NPM, IC	Unit subgroups	Open text response	A100	
STUDID	Consortium	Student ID	00001 to 99999	F5.0	
PASSWORD	Consortium	Test system password	Alphanumeric	A6	
GENDER	NPM, IC	Student gender	1=Male 2=Female	F1.0	7=N/A 8=Invalid 9=Missing
STUDAPT	NPM, IC	Student aptitude measure	Open text response	A100	
BROADFIELD	NPM, IC	Broad field of study	ISCED code (see Annex)	F1.0	
NARROWFIELD	NPM, IC	Narrow field of study	ISCED code (see Annex)	F2.0	
DOB	NPM, IC	Student date of birth	YYYYMMDD	F8.0	7=N/A 8=Invalid 9=Missing
STATUS	NPM, IC	Student eligibility status	1=Included 2=Exempted 3=Excluded	F1.0	7=N/A 8=Invalid 9=Missing

*Step 6: Selecting the third-stage sample*

The NPM will select a random sample of 10 students from the list provided by the IC by:

- Adapting the script given in Figure 6 and drawing the sample of 10 students per UNIT3;
- Recording and saving the log of the sample selection runs (one log for each strand and each institution, clearly named e.g. <INSTID>STD\_SAMPLE);
- For quality assurance purposes, uploading a copy of working version of the frame, the log (including the random seed used) and the resulting sample to your country's secure space on the AHELO Exchange; and
- Storing copies of the logs and samples for later reference and sending copies of the selected samples to the respective ICs.

```

GET FILE= 'DRIVE:\FOLDER\<INSTID>STD_WFRAME.SAV' .

SORT CASES BY UNIT1 (A) UNIT2 (A) UNIT3 (A) .

CSPLAN SAMPLE
  /PLAN FILE= 'DRIVE:\FOLDER\<INSTID>STD.csplan'
  /PLANVARS SAMPLEWEIGHT=SampleWeight_Final_3
  /PRINT PLAN
  /DESIGN STAGELABEL='third'
  STRATA=UNIT1 UNIT2 UNIT3
  /METHOD TYPE=SIMPLE WOR
    
```



```

ESTIMATION=DEFAULT
/SIZE VALUE=10
/STAGEVARS INCLPROB (InclusionProbability_3)
CUMWEIGHT (CumulativeWeight)
POPSIZE (PopulationSize_3)
SAMPSIZE (SampleSize_3)
WEIGHT (SampleWeight_3) .

CSSELECT
/PLAN FILE= 'DRIVE: \FOLDER\<INSTID>STD.csplan'
/CRITERIA STAGES=1
SEED=123454321
/CLASSMISSING EXCLUDE
/DATA RENAMEVARS
/SAMPLEFILE OUTFILE= 'DRIVE: \FOLDER\<INSTID>STD_SAMPLE.sav'
/PRINT SELECTION.

```

**Figure 7: SPSS syntax for the selection of a (stratified) simple random sample of 10 students per selected UNIT3 (course section) in each selected UNIT1 (department or similar)**

The following remarks are made regarding the second sampling approach:

- Multi-stage sampling of students need not be the most statistically efficient method to draw a sample of students, but it may be the only practical method to complete that task in some institutions. NPMs and ICs should try their best to implement the single-stage sampling strategy in each of their participating institutions;
- Estimates from the multi-stage sample of students – as described here – may suffer from undercoverage bias as finishing students who only take electives cannot be sampled;
- Stages 1 and 2 could be combined into a single stage (selecting a random sample of 20 sections or groups throughout the institution) if a comprehensive and exhaustive list of course sections of interest existed;
- Alternatively, stages 2 and 3 could be combined into a single stage (selecting a random sample of 20 students from each selected department) if an institution-level comprehensive list of in-scope of students did not exist but did exist within each department, and lack of time or resources made it impossible to gather all lists from all departments, unduplicate them and assemble them into a single institution-level list frame of students;
- If a multi-stage of students appears to be the only feasible way to obtain random samples of students, the sample of faculty can be limited to those 10 randomly selected departments, selecting 4 faculty per selected department;
- In case of doubt, NPMs and ICs are urged to contact the Sampling Team ([ahelo@acer.edu.au](mailto:ahelo@acer.edu.au)) for support and advice as often as required.

## Annex: ISCED Broad groups and fields of education

For reference, classification codes for the International Standard Classification of Education (ISCED) are given in Table 7 (broad groups, for the AHELO variable BROADFIELD) and Table 8 (fields of education, for the AHELO variable NARROWFIELD). These codes are copied from UNESCO (2006).

**Table 7: ISCED broad groups of education (BROADFIELD)**

Variable	Value
General Programmes	0
Education	1
Humanities and Arts	2
Social sciences, business and law	3
Science	4
Engineering, manufacturing and construction	5
Agriculture	6
Health and welfare	7
Services	8

**Table 8: ISCED fields of education (NARROWFIELD)**

Variable	Value	Description
Basic programmes	01	Basic general programmes pre-primary, elementary, primary, secondary, etc.
Literacy and numeracy	08	Simple and functional literacy, numeracy.
Personal development	09	Enhancing personal skills, e.g. behavioural capacities, mental skills, personal organizational capacities, life orientation programmes.
Teacher training and education science	14	Teacher training for pre-school, kindergarten, elementary school, vocational, practical, non-vocational subject, adult education, teacher trainers and for handicapped children. General and specialized teacher training programmes. Education science: curriculum development in non-vocational and vocational subjects. Educational assessment, testing and measurement, educational research, other education science.
Arts	21	Fine arts: drawing, painting, sculpture; Performing arts: music, drama, dance, circus; Graphic and audio-visual arts: photography, cinematography, music production, radio and TV production, printing and publishing; Design; Craft skills.
Humanities	22	Religion and theology; Foreign languages and cultures: living or 'dead' languages and their literature, area studies; Native languages: current or vernacular language and its literature; Other humanities: interpretation and translation, linguistics, comparative literature, history, archaeology, philosophy, ethics.
Social and behavioural science	31	Economics, economic history, political science, sociology, demography, anthropology (except physical anthropology), ethnology, futurology, psychology, geography (except physical geography), peace and conflict studies, human rights.
Journalism and information	32	Journalism; library technician and science; technicians in museums and similar repositories; Documentation techniques; Archival sciences.
Business and administration	34	Retailing, marketing, sales, public relations, real estate; Finance, banking, insurance, investment analysis; Accounting, auditing, bookkeeping; Management, public administration, institutional administration, personnel administration; Secretarial and office work.

Variable	Value	Description
Law	38	Local magistrates, 'notaires', law (general, international, labour, maritime, etc.), jurisprudence, history of law.
Life sciences	42	Biology, botany, bacteriology, toxicology, microbiology, zoology, entomology, ornithology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences.
Physical sciences	44	Astronomy and space sciences, physics, other allied subjects, chemistry, other allied subjects, geology, geophysics, mineralogy, physical anthropology, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, marine science, vulcanology, palaeoecology.
Mathematics and statistics	46	Mathematics, operations research, numerical analysis, actuarial science, statistics and other allied fields.
Computing	48	Computer sciences: system design, computer programming, data processing, networks, operating systems - software development only (hardware development should be classified with the engineering fields).
Engineering and engineering trades	52	Engineering drawing, mechanics, metal work, electricity, electronics, telecommunications, energy and chemical engineering, vehicle maintenance, surveying.
Manufacturing and processing	54	Food and drink processing, textiles, clothes, footwear, leather, materials (wood, paper, plastic, glass, etc.), mining and extraction.
Architecture and building	58	
Agriculture, forestry and fishery	62	Agriculture, crop and livestock production, agronomy, animal husbandry, horticulture and gardening, forestry and forest product techniques, natural parks, wildlife, fisheries, fishery science and technology.
Veterinary	64	Veterinary medicine, veterinary assisting.
Health	72	Medicine: anatomy, epidemiology, cytology, physiology, immunology and immunohaematology, pathology, anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, neurology, psychiatry, radiology, ophthalmology; Medical services: public health services, hygiene, pharmacy, pharmacology, therapeutics, rehabilitation, prosthetics, optometry, nutrition; Nursing: basic nursing, midwifery; Dental services: dental assisting, dental hygienist, dental laboratory technician, odontology.
Social services	76	Social care: care of the disabled, child care, youth services, gerontological services; Social work: counselling, welfare n.e.c.
Personal services	81	Hotel and catering, travel and tourism, sports and leisure, hairdressing, beauty treatment and other personal services: cleaning, laundry, dry-cleaning, cosmetic services, domestic science.
Transport services	84	Seamanship, ship's officer, nautical science, air crew, air traffic control, railway operations, road motor vehicle operations, postal service.
Environmental protection	85	Environmental conservation, control and protection, air and water pollution control, labour protection and security.
Security services	86	Protection of property and persons: police work and related law enforcement, criminology, fire-protection and fire fighting, civil security; Military.