Table of Contents

AMERICAN
INSTITUTES
FOR RESEARCH

AMERICAN
INSTITUTES
FOR RESEARCH

AMERICAN
INSTITUTES
FOR RESEARCH

1. INTRODUCTION

1.1. Overview

Over the past several years the OECD, and Network C in particular, have shown increasing interest in understanding school processes in their own right. This is reflected by a cycle of school surveys that began in 1995 at the primary level and extends now to the International Survey of Schools at the Upper Secondary Level (USSS). These school surveys are designed to produce institutional process indicators that reveal some of the key characteristics in both the learning environment and the organization of schools that may contribute to better student achievement.

Surveys are an important tool in Network C's mission of focusing policy makers' attention on processes. Responding to the Governing Board's request, Network C has begun a systematic series of surveys that will lead to a rich source of data to be presented as indicators in *Education at a Glance*. To further its goal of enlightening the processes by which schools educate children, Network C is now launching a second school survey focusing on the upper secondary level, which the American Institutes for Research (AIR) will lead as the International Study Coordinator. After reviewing a series of issues through thematic papers written by experts, Network C selected four topic areas for this survey:

- School characteristics aimed at facilitating the transition to the labor market and/or further education;

- Conditions of schooling that are instrumental in enhancing educational quality;

- Human resources; and

- Availability and use of information and communication technology (ICT).

These four areas have been chosen because they demonstrate school organizational processes that have a direct impact on the quality of learning.

A major objective of any international comparative study is the selection of quality samples since only properly selected samples will yield unbiased, accurate, and internationally comparable results. For the USSS to yield data that AIR can use in its analyses, it is paramount that this manual be referred to and followed throughout the sampling process.

The procedures outlined in this and other USSS manuals have been designed to ensure that high quality, internationally comparable data will be available for analyses. If you deviate from these procedures without the prior approval of the American Institutes for Research, you risk losing

AMERICAN
INSTITUTES
FOR RESEARCH

the ability to compare your country's results on USSS with those of the other countries. It is extremely important that the procedures outlined in this and the other USSS manuals are followed. Any anticipated deviation from these procedures should be brought to the attention of the technical staff at AIR prior to implementation.

► Do not modify any procedures without consulting AIR.

This manual contains a series of forms that you, as National Project Manager (NPM), must complete as you carry out the USSS sampling procedures. These forms require the vital information that is necessary to document the sampling activities and to ensure that the national school samples meet all of the AIR quality standards. You should submit the forms to Eugene Johnson at AIR, as is indicated on the first page of Appendix A.

USSS demands hard work and expertise from everyone involved, but especially from National Project Managers. One consistent finding from previous AIR studies is that the National Centers that follow study procedures produce the highest quality data with the least amount of effort. Your contribution as National Project Manager for your country is crucial to the current and future success of this survey.

Please keep a record of all problems and unforeseen situations. You are requested to document your experiences—the good and the bad—along with suggestions for further improvements in the *Survey Activities Report* (see chapter 7 of the *USSS Survey Operations Manual* Ref. No. 01-01-01). Your input will be invaluable to our efforts to improve future studies.

1.2.   Important Deadlines for USSS

It is critical that all tasks related to sampling be carried out in accordance with the international schedule so that the data are collected and ready for analysis in a timely manner. If you are at risk of missing any of the international deadlines, you should contact AIR immediately. Table 1.1 presents some important dates for the International Survey of Schools at the Upper Secondary Level. If you cannot meet the international schedule for sampling, survey administration, and data delivery, you risk having your data omitted from the analyses.

You are expected to have all samples approved by AIR by **May 1, 2001**.

► If your country is participating in the Field Test, you are expected to identify the Field Test sample and secure cooperation by **March 15, 2001**. Please see the timeline below for additional information.

AMERICAN
INSTITUTES
FOR RESEARCH

Table 1.1: Important Dates for the International Survey of Schools at the Upper Secondary Level

NOTE: All activities specifically related to sampling procedures have been underscored.

Month and Day, 2001		Activities
January	**15**	*Materials* Receive sampling plan for main study Receive guidelines for implementing Field Test (only participating countries)
February	**15**	Receive revised questionnaires incorporating country Recommendation for pre-piloting *Materials* Receive *USSS Sampling Manual* Receive *USSS Survey Operations Manual*
March	**1**	Receive revised questionnaire for Field Test translation
	15	Sampling frame sent to AIR for approval (with forms) Secure cooperation of Field Test sample
April	**1**	Translated Field Test questionnaire to AIR for approval Draft sampling plan to AIR (with forms)
	15	Begin Field Test
May	**1**	*Materials* Receive data entry database (with instructions)
	15	Send drawn sample to AIR for approval
	31	End Field test
June	**1**	Samples approved by AIR
	15	AIR receives all coded Field Test data from countries
August	**1**	*Materials* Receive final revised questionnaire
September	**1**	Translations sent to AIR for approval
	15	AIR releases report on Field Test results
October	**1**	Begin main study data collection
December	**23**	Send all coded data back to AIR Send the *Survey Activities Report* to AIR no later than two weeks after submitting your coded data.

AMERICAN
INSTITUTES
FOR RESEARCH

Month and Day, 2001	Activities
Month and Day, 2002	Activity
May 15	AIR releases final report on main study

1.3. The School Sampling Manual

1.3.1. Purpose of this Manual

The *International Survey of Schools at the Upper Secondary Level School Sampling Manual* has been prepared by the staff of the American Institutes for Research. It is designed to provide the National Project Manager and other National Center staff members of each country with the school sampling procedures for the USSS Field Test[1] and Main Survey. It specifies the timelines, actions, and procedures that must be followed to develop and implement the USSS school sample design.

This manual serves as a procedures manual to carry out all of the required school sampling activities. NPMs can count on AIR to be an invaluable resource in assisting them in the development and implementation of their national sampling plans.

It remains the responsibility of NPMs to ensure that their national sample design is approved by AIR and is fully implemented. AIR will assist NPMs in meeting these two critical objectives.

1.3.2. Audience and Use of the Manual

This manual is written for you, the National Project Manager, and for your National Center colleagues who are responsible for assisting you with the sampling procedures. The *USSS School Sampling Manual* will guide you through the various sampling phases leading to the selection of the national school sample. This includes the initial planning, constraint recognition, sample selection procedures, and documentation of the national sampling plan.

This *International Survey of Schools at the Upper Secondary Level School Sampling Manual* is a very important resource for you. It is essential that you read this manual carefully.
For all sampling activities, NPMs are required to complete and submit sampling forms to AIR. These sampling forms are described throughout this manual and are presented in Appendix A. These forms are used to document all sampling activities and to ensure that national school samples meet all of the AIR quality standards.

---

[1] Not all countries will be participating in the Field Test.

AMERICAN
INSTITUTES
FOR RESEARCH

► Please do not hesitate to contact AIR for clarification if anything in this manual is unclear.

► National Project Managers should refer to this manual for detailed information on the sampling methodology as well as specific details and worksheets to select their country's sample.

1.3.3. Reasons for the Sampling Procedures

The sampling processes outlined in this and other American Institutes for Research manuals have been designed to ensure that high quality, internationally comparable data will be available for analyses. If you deviate from these procedures without the prior approval of AIR, you risk losing the ability to compare your country's results on the International Survey of Schools at the Upper Secondary Level with those of the other countries. It is extremely important that the procedures outlined in this and other AIR manuals are followed. Any anticipated deviation from these procedures should be brought to the attention of the technical staff of AIR prior to implementation.

1.3.4. Notes on the Field Test

Prior to the main administration of the OECD Upper Secondary School Survey, several countries will field test the instrument and the procedures for its administration. The goal of the Field Test is not to obtain questionnaire results that represent your country, but rather to make sure that the instrument works as expected and that any problems with field administration, response rates, etc. can be corrected before the main administration.

► If your country is participating in the Field Test, please read the remainder of this section. If not, proceed to Section 1.3.5.

The instrument will be field tested in 32 schools per country. To keep the cost of administration down, we do not expect countries to draw a representative sample of school sites for the Field Test. Instead, the sample will be purposive (that is, designed to cover a variety of different schools). We expect countries to obtain variation across the school sites they select along the following variables:

- Predominance of program type in school site (i.e., 3A, 3B, and 3C);

- Size of school site;

- Aggregate socioeconomic status of school site; and

- Location of school site (urban vs. rural)

Additionally, it is desirable to obtain representation across other key characteristics that might be expected to produce different results or challenges to response and which define major subgroups of schools. Countries that participate in the Field Test will be expected to carry out the same activities as they would in the main administration, namely:

- Contacting school sites to secure participation;

- Tracking and reporting the response rate;

- Translating the instrument;

- Sending the instrument;

- Implementing follow-up procedures until the instrument is returned;

- Entering questionnaire data into the database;

- Writing a brief report documenting Field Test procedures and any problems encountered; and

- Sending the Field Test database and the brief report to AIR for analysis.

These steps are documented in the *USSS Survey Operations Manual* Ref. No. 01-01-01.

The Field Test will be conducted between **April** and **May 2001**, with all Field Test data files and materials due to AIR on **June 15, 2001**.

► All Field Test data files and materials are due to AIR **NO LATER THAN June 15, 2001**.

AMERICAN
INSTITUTES
FOR RESEARCH

1.3.5. Organization of the Manual

This manual provides detailed information for each phase of the school sample design. The chapters following this introductory section are organized into a series of five sampling steps. By following these steps, NPMs will be able to design their national sampling plan and select school samples that will meet the AIR quality standards.

Chapter 2 describes the procedures for the selection of the school sample.

Chapter 3 provides Step 1 of the sampling process — defining the target population.

Chapter 4 outlines Step 2 — stratifying the sample.

Chapter 5 details Step 3 — developing a national sampling plan.

Chapter 6 outlines Step 4 — constructing the school sampling frame.

Chapter 7 provides Step 5 — selecting the school sample.

Chapter 8 provides Step 6 — computing and checking sampling weights.

In addition, two Appendices are also included in this manual:

- **Appendix A** contains copies of all Sampling Forms needed to document the sampling procedures:

Sampling Form 1:	*National Desired Target Population*
Sampling Form 2:	*National Defined Target Population*
Sampling Form 3:	*Stratification*
Sampling Form 4:	*Small Schools*
Sampling Form 5:	*Sampling Frame Description*
Sampling Form 6:	*School Tracking Form*
Sampling Form 7:	*Excluded Schools*
Sampling Form 8:	*Population Counts by Strata*
Sampling Form 9:	*Sample Allocation by Explicit Strata*
Sampling Form 10:	*School Sample Selection*
Sampling Form 11:	*School Sampling Frame*
Sampling Form 12:	*Response Rate*
Sampling Form 13:	*Calculating the School Weights*

- **Appendix B** provides the *Sample Size Table*. You will need to use this to determine the actual number of schools to be sampled in your country.

AMERICAN
INSTITUTES
FOR RESEARCH

1.4. Important Contact Information

The American Institutes for Research, along with its network of consultants and experts, will direct the International Survey of Schools at the Upper Secondary Level. General questions about the project may be directed to either Dr. Joel Sherman or Dr. Stéphane Baldi. <u>Questions about sampling and survey operations should be directed to either Dr. Eugene Johnson or Dr. Stéphane Baldi.</u> Questions about data entry should be directed to Mr. Steven Honegger, and questions about data analysis should be directed to Dr. Marianne Perie. Detailed contact information is provided below. If you are in doubt as to which individual to contact with a specific question, please contact Dr. Joel Sherman for advice.

The American Institutes for Research – Dr. Joel Sherman, Project Director
Pelavin Research Center
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835
UNITED STATES
Telephone: +1 (202) 944-5361 / Fax: +1 (202) 944-5454
e-mail: jsherman@air.org

The American Institutes for Research – Dr. Stéphane Baldi, Deputy Project Director & <u>Sampling and Survey Operations</u>
Pelavin Research Center
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835
UNITED STATES
Telephone: +1 (202) 298-2615 / Fax: +1 (202) 944-5454
e-mail: sbaldi@air.org

The American Institutes for Research – Dr. Eugene Johnson, <u>Sampling and Survey Operations</u>
Pelavin Research Center
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835
UNITED STATES
Telephone: +1 (202) 944-5433 / Fax: +1 (202) 944-5454
e-mail: ejohnson@air.org

AMERICAN
INSTITUTES
FOR RESEARCH

The American Institutes for Research – Mr. Steven Honegger, Database Development and Data Entry
Pelavin Research Center
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835
UNITED STATES
Telephone: +1 (202) 944-5393 / Fax: +1 (202) 944-5454
e-mail: shonegger@air.org

The American Institutes for Research – Dr. Marianne Perie, Data Analysis and Reporting
Pelavin Research Center
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835
UNITED STATES
Telephone: +1 (202) 944-5388 / Fax: +1 (202) 944-5454
e-mail: mperie@air.org

Network C Secretariat – Dr. Jaap Scheerens
University of Twente
Faculty of Educational Science and Technology
Department of Educational Organisation and Management
P.O. Box 217
7500 AE Enschede
THE NETHERLANDS
Telephone: +31 (53) 489-4579 / Fax: +31 (53) 489-3791
e-mail: scheerens@edte.utwente.nl

Organisation for Economic Co-operation and Development – Dr. Claudia Tamassia
2, Rue André-Pascal
75775 Paris Cedex 16
FRANCE
Telephone: +33 (1) 45-24-1903 / Fax: +33 (1) 45-24-9098
e-mail: Claudia.TAMASSIA@oecd.org

2. OVERVIEW OF SAMPLING

2.1. Introduction

The OECD International Survey of Schools at the Upper Secondary Level sampling plan has been developed by the American Institutes for Research following guidelines set forth by the OECD and Network C. These guidelines are that 1) the samples drawn by countries allow the production of indicators framed in terms of the percentage of students in schools with certain characteristics, and 2) the samples reflect the major national program orientation so as to allow estimates at the program level. AIR has developed a one-stage sample of school sites drawn probability proportional to size (PPS) that meets the guidelines set forth above while greatly simplifying the operational aspects of implementing the sample design. This chapter has three primary objectives:

- To specify the sampling responsibilities of both AIR and participating countries;

- To describe the target population, stratification of the sample, and the selection of substitute school sites; and

- To provide the data requirements for assembling the sampling frame.

2.2. Sampling Responsibilities

2.2.1. Sampling Responsibilities of AIR

AIR's responsibilities center around three key areas of sampling:

1) Design the sampling plan
   - Determine the most appropriate sampling design for USSS and specify the following specific features of the sampling plan:
     - i. Number of sampling stages and the sampling units in each of the stages;
     - ii. Stratification of the sample and potential stratification variables;
     - iii. Sampling strategy within each stratum, (e.g., Simple Random Sample, Probability Proportional to Size); and
     - iv. Sample size based on statistical precision requirements.

2) Provide technical assistance and approve samples
   - Work with countries in identifying target population;
   - Offer on-going consultation for sampling operations;

AMERICAN
INSTITUTES
FOR RESEARCH

- Review and approve each country's sampling frame prior to drawing the sample; and
- Review and approve each country's sample based on data and documentation provided by countries.

3) Develop a sampling manual

- Prepare a sampling manual to provide step-by-step guidance to participating countries on how to implement the sampling plan. The sampling manual will give instructions on:
    i. How to prepare a sampling frame that best reflects the targeted population;
    ii. How to stratify the sample and allocate the sample to each stratum;
    iii. How to implement the sampling strategy within each stratum;
    iv. How to calculate the sampling probabilities and sampling weights;
    v. How to verify the accuracy of the sampling and weighting procedures;

2.2.2. Sampling Responsibilities of Participating Countries

Participating countries have responsibilities in three important areas of sampling:

1) Prepare a sampling frame

- Based on the instructions provided by AIR, each participating country needs to prepare a sampling frame that best reflects the targeted population for the study. This sampling frame should include, at minimum, the following information for the first stage of selection:
    i. The name of each eligible school site;
    ii. The contacting address for each school site;
    iii. If stratification is involved in the sampling, information on stratification variables should be available for each of the school sites (e.g., urban/rural; region of the country; etc.); and
    iv. Enrollment size at each of the school sites for all in-scope programs (i.e., ISCED 3), and preferably by program.

- Countries are expected to give AIR the data file they plan to use as a frame for review and approval prior to drawing the sample.

2) Implement the sampling plan

- Following the instructions provided in the sampling manual, each participating country is required to implement the sampling plan in its own country. Specifically, each country needs to do the following:
    i. Prepare an appropriate sampling frame and put it into an electronic data set;
    ii. Stratify the sampling frame and appropriately allocate the sample to various strata;
    iii. Implement the probability sampling within each stratum;
    iv. Calculate the sampling probabilities and sampling weights;

  v. Check the representativeness of the selected sample and the correctness of the weighting procedure; and

  vi. After the survey administration, calculate the response rates and derive the final sampling weights that take into account non-response rates.

- Once the sample is drawn, countries are expected to send it with accompanying documentation (e.g., computer programs) to AIR for review and approval.

3) Document and report the sampling process

- Each participating country needs to carefully document every step of the sampling process and to produce a final report. The final report should be sent to AIR to facilitate its quality control and data verification function. The final report should include the following information:

  i. The sampling frame used in the process and a basic assessment about the appropriateness of this sampling frame;

  ii. The stratification involved in the sampling, and the reasons for the stratification.

  iii. Summary statistics concerning upper secondary education, including number of school sites, enrollment, number of teachers at both national level and stratum level;

  iv. Documentation of adherence to the procedures developed by AIR to calculate the sampling probabilities;

  v. Documentation of adherence to the procedure developed by AIR to calculate the response rates and the final sampling weights; and

  vi. An electronic copy of the selected sample of school sites with all of the following information: enrollment, program type, information concerning stratification variables, selection probabilities, and sampling weights.

2.3. Target Population

The target population for the 2001 Upper Secondary School Survey sample is the set of all school sites containing at least one upper secondary program (as defined by Phase I of the study). In the vast majority of countries, a school site will be equivalent to a school. However, in countries where a school encompasses many educational units, the school site will be the physical entity where the upper secondary program is administered.

The USSS sample will consist of up to 400 school sites, selected with probability proportional to the site's enrollment of students in upper secondary programs.[2] The number of school sites was

---

[2] This sample size assumes an infinite number of school sites in a country. Hence, the sample size will be reduced based on the actual number of school sites in a country (formulas are provided in Chapter 8 and a table of required sample sizes is provided in Appendix B).

selected to meet a sampling precision requirement equivalent to that of other well-known international studies. This requirement was designed to ensure that each country draws a large enough sample so that, with 95 percent probability, the sample estimates of population values for means will be within +/- 0.1 of the standard deviation, correlations will be within +/- 0.1, and for proportions will be within +/- 5.0 percent.

2.4. Stratification

Prior to sampling, school sites in the sampling frame can be assigned to a predetermined number of strata. This stratification may be explicit or implicit. There are basically three reasons for stratifying:

- To produce reliable estimates for sub-national domains;

- To improve the sampling efficiency, thereby improving the reliability of national estimates; and

- To ensure that different parts of the population are appropriately represented in the sample.

Prior to sampling, schools can be ordered, or stratified, in the sampling frame. Stratification consists of grouping schools into strata according to some grouping, or stratification variables. Examples of commonly used stratification variables include:

- Regions (states, provinces)

- Urbanization (rural, urban)

- Socioeconomic status

- School types (public, private)

Countries will have the option to use a combination of explicit or implicit stratification variables in drawing their dataset, thus allowing them to increase precision in representing specific groups of interests (e.g., private schools). Explicit strata are distinct subsets of the sampling frame within which a specified number of schools are selected. Implicit strata are groupings of schools within explicit strata designed to ensure a proportional allocation across the implicit strata.

AMERICAN
INSTITUTES
FOR RESEARCH

2.5. Selecting Substitute School Sites

As the school sites are sampled, replacement sites are simultaneously identified should they be needed to replace non-participating sampled sites.

> ► The required response rate for this study is 50% of initially sampled school sites, and 85% after replacement schools are used to avoid flagging of your country's results.

Each sampled school site in the survey will be assigned two replacement school sites in the sampling frame, inasmuch as it is possible given the constraints imposed by the sampling guidelines. These constraints state that a sampled school site cannot be designated as a replacement school site, and a given school site cannot be assigned to substitute for more than one sampled school site.

For each sampled school site, the next two school sites immediately following it in the sampling frame will be designated as its replacement school sites. The use of implicit stratification variables and the subsequent ordering of the school site sampling frame by size ensure that any sampled school site's replacements will have similar characteristics.

When a sampled school site is the last school listed in an explicit stratum, then the two school sites immediately above it will be designated as its replacement school sites. If a sampled school site is the next to last school site listed in an explicit stratum, then the school sites immediately above and below it will be designated as its replacement school sites.

2.6. Data Requirements for Assembling the Sampling Frame

As a first step in drawing the sample, you must assemble a sampling frame that contains the universe for the population to be sampled. The data requirements for assembling this sampling frame are of two kinds: mandatory requirements and optional requirements.

2.6.1. Mandatory Requirements

Countries must obtain the following information in order to construct their sampling frame for the USSS. This information must be forwarded (in electronic format) to both Eugene Johnson and Stéphane Baldi for review and approval **NO LATER THAN March 15, 2001**.

- A list of all school sites in the country offering at least one upper secondary program (as defined by Phase I of the study);

- For each of these school sites, a unique numerical ID number (typically assigned by the national government);

- The complete address of each school site, including a contact name and phone number; and

- Total enrollment in all in-scope upper secondary programs at that site. (If this information is not available, then other suitable measures of size could be the total enrollment at the school site or the average enrollment in target grades).

2.6.2. Optional Requirements:

In addition, if countries decide to stratify their sample for greater precision or specific reporting categories, as discussed in Section 2.3, they will need measures of their stratification variables. These could include:

- Regions (states, provinces)

- Urbanization (rural, urban, suburban)

- Socioeconomic status at the school site level (e.g., % of students in poverty)

- School types (public, private)

- Percentage of students with certain characteristics (e.g., race, etc.)

► All information related to the construction of the sampling frame must be forwarded to AIR in electronic format by **March 15, 2001**.

2.7. Reminder: Resources and Documentation

As a reminder, questions about sampling should be directed to either Dr. Eugene Johnson or Dr. Stéphane Baldi whose contact information is listed in Section 1.4.

The *International Survey of Schools at the Upper Secondary Level School Sampling Manual* is part of a set of materials designed to provide you with the information you need to conduct USSS successfully. The other important materials are:

- *USSS Survey Operations Manual*

- USSS Data Entry Software and Documentation

AMERICAN INSTITUTES FOR RESEARCH

► If you are missing any of these materials, you should immediately contact AIR.
  REMINDER: Date Entry Software will be provided to NPMs on **May 1, 2001.**

3.   STEP 1 — TARGET POPULATION DEFINITION

3.1. Overview

This chapter has one primary objective:

- To clarify the relationships among the levels of target populations and the possible exclusions leading to reductions in coverage.

3.2. International Desired Target Population

The international desired target population for the 2001 USSS sample is the set of all school sites containing at least one upper secondary program (as defined by Phase I of the study). In the vast majority of countries, a school site will be equivalent to a school. However, in countries where a school encompasses many educational units, the school site will be the physical entity where the upper secondary program is administered.[3]

All NPMs must determine the appropriate target schools in accordance with the above definition.

In simple terms, the international desired target population is intended to provide full coverage of all eligible schools in a country's education system. Any deviation from full national coverage needs to be described and quantified for publication in the international reports. Any questions regarding the identification of the target population should be directed to Dr. Eugene Johnson (see Section 1.4 for contact information).

3.3. National Desired Target Population

National Project Managers must define and describe the national desired target population for their country. This definition must address the requirements of the international desired target population. It should provide the most exhaustive national coverage of eligible schools possible.

The national desired target population must be described on *Sampling Form 1*. In cases where this population deviates from full national coverage of all eligible schools, the deviations must be described and school data provided to measure the extent of the reduced coverage. Any

---

[3] In the remainder of this manual we will use school to mean either school or school site depending on a country's situation.

deviations, in terms of reduced coverage, from the international desired target population will be presented in the international reports.

AIR encourages all NPMs to provide complete national coverage in their national desired target population. Political, organizational, or operational reasons, however, might make it extremely difficult for some NPMs to meet this objective. Consequently, when defining their national desired target population, NPMs may choose to reduce their national coverage. This could be done by removing a geographical region, an educational sub-system, or even a language group.

In this case, the national desired target population will differ from the international desired target population. Significant differences in terms of the number of schools excluded from the national desired target population will mean that survey results will not be deemed representative of the entire national schools system.

► National Project Managers are strongly encouraged to provide <u>complete</u> national coverage when defining their national desired target population. Limitations in national coverage will be annotated in the international reports.

By defining the national desired target population in *Sampling Form 1* and describing any limitations in national coverage, this information will be properly documented and can be formally presented in the international reports.

3.4. National Defined Target Population

Using their national desired target population as a general framework, NPMs must describe their national defined target population. The national defined target population is essentially the sampling frame. Differences between the national desired and defined target populations will generally arise out of practical considerations and should be kept to a minimum. All exclusions must not exceed 5% of the national desired target population. Countries who exceed this limit will be annotated in the international reports.

► NPMs must keep exclusions to a minimum: no more than 5% of the national desired target population. Large exclusion rates will be annotated in the international reports.

3.5. Exclusions

The national defined target population may represent a subset of the national desired target population. All schools from the national desired target population excluded from the national defined target population are referred to as the excluded population.

One reason NPMs might exclude schools is because they offer a curriculum, or school structure, radically different from the mainstream educational system.

Other reasons might be considered, but NPMs should verify the reasons with AIR before excluding. All such exclusions must be described and quantified in *Sampling Form 2*.

Note: The target population is limited to schools that contain at least one upper secondary program. Schools that do not offer this program, such as elementary schools, are therefore not considered part of the excluded population.

Example 1	**National Defined Target Population**

The following is an example of a national defined target population. The fictitious example uses United States school data.

In this example, the United States has chosen to exclude all U.S. special schools enrolling students with disabilities because their curriculum is atypical when compared to the rest of the country. These school exclusions account for 3.5% of eligible schools in the national desired target population.

3.6. Coverage and Exclusions

The distinctions among the international desired, national desired, and national defined target populations can be nebulous. This study aims to achieve full coverage of the international desired target population among all participating countries and to keep all exclusions to a minimum.

The difference between the international and national desired target populations is labeled as "exclusions from national coverage." As a guiding principle, only a sizeable exclusion of the target population, whereby the remainder can be succinctly described and in a coherent manner, would be considered. Countries will need to consult AIR to clarify their situation in regards to coverage and exclusions.

AMERICAN
INSTITUTES
FOR RESEARCH

All other sources of exclusions would constitute exclusions from the national desired target population. Exclusions therefore describe the difference between the national desired and defined target populations. The following chart illustrates the relationships among the many levels of target populations and the possible reductions in coverage and exclusions.

```
                        ┌─────────────────────────────────────────┐
                        │  International Desired Target Population  │
                        └─────────────────────────────────────────┘
                            ↙                              ↘
    ┌──────────────────────────────────┐    ┌──────────────────────────────────┐
    │ National Desired Target Population│    │  Exclusions from National Coverage│
    └──────────────────────────────────┘    └──────────────────────────────────┘
         ↙                    ↘
┌──────────────────────────────────┐    ┌──────────────────────────────┐
│ National Defined Target Population│    │    School-Level Exclusions   │
└──────────────────────────────────┘    └──────────────────────────────┘
```
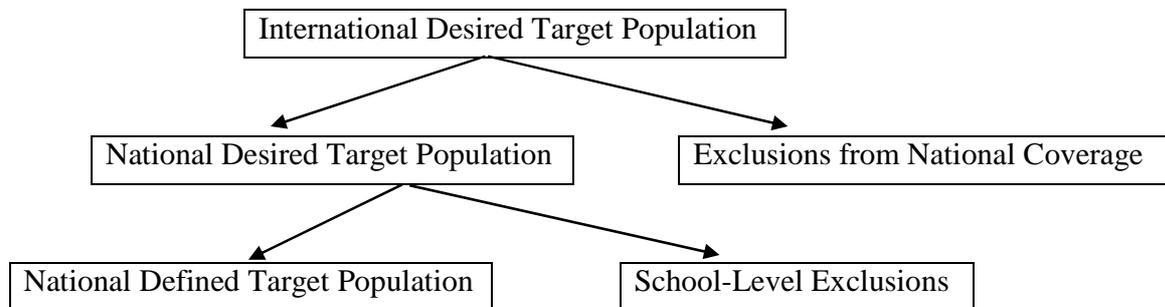
4.   STEP 2 — STRATIFICATION

4.1. Overview

Prior to sampling, schools can be grouped, or ordered, in the sampling frame. This is called stratification. Stratification consists of grouping schools into strata or ordering schools within strata according to some stratification variables. Stratification is generally used for the following reasons:

- To improve the efficiency of the sample design, thereby making survey estimates more reliable;

- To apply different sample designs, such as disproportionate sample allocations, to specific groups of schools, such as schools in different states and provinces; and

- To ensure proportional representation of specific groups of the target population in the sample.

Examples of stratification variables include:

- Regions (states, provinces);

- Urbanization (rural areas, urban areas);

- Socioeconomic status (low, medium, or high income); and

- School types (public, private).

Stratification can be either explicit or implicit.

4.2. Explicit Stratification

Explicit stratification consists of building separate school lists, or sampling frames, according to one or more variables. For example, if geographic region is an explicit stratification variable, then separate school sampling frames would be constructed for each geographic region. Possibly different sample designs, or simply a constant sample size, would then be applied to each school sampling frame to select the sample of schools.

In this survey, the major reason for considering explicit stratification is to implement a disproportionate allocation of the school sample to the explicit strata. For example, the same

AMERICAN
INSTITUTES
FOR RESEARCH

number of schools could be sampled from each explicit stratum, regardless of the relative size of each stratum. The objective in this situation would be to produce equally reliable estimates for each stratum.

Another important reason for defining explicit strata is to deal with specific sample design issues. For example, explicit strata might be required to deal with small schools, as well as with very large schools. Small schools are discussed in greater detail in Sections 5.7 and 6.8. Very large schools are discussed in Section 7.4.

4.3. Implicit Stratification

Implicit stratification consists of sorting the school sampling frame by a set of implicit stratification variables. This type of stratification is very effective for the school sample selection method proposed for USSS. It is a very simple way of ensuring a strictly proportional sample allocation of schools across all implicit strata. It can also lead to improved reliability of survey estimates, provided the implicit stratification variables being considered are known to have a significant between-strata variance component.

4.4. Stratification Guidelines

The selection of two to four stratification variables is usually sufficient. The following guidelines should be applied when selecting stratification variables:

- Each school in the sampling frame should belong to one and only one explicit stratum;

- Good measures of size should be known for the explicit strata. In particular, for each explicit stratum, both the number of schools and their enrollment should be known. Normally, these should be easily computed directly from the school sampling frame.

- Defining very small strata, especially explicit strata, should be avoided because this is unlikely to improve the overall level of sampling precision. A few divisions of a continuous stratification variable usually provide nearly all of the gains in sampling precision available from that variable.

- At least two schools should be selected in each explicit stratum to permit the computation of sampling error. This criterion usually sets an upper limit to the number of strata that can be defined.

AMERICAN
INSTITUTES
FOR RESEARCH

NPMs must describe all variables they intend to use for stratification, both explicit and implicit, in *Sampling Form 3*. They should also indicate the total number of strata that will be defined.

Example 2   **Stratification**

The following is an example of stratification.

Country X defined two stratification variables:

1. **State** – The eight levels of this variable correspond to the eight states in Country X. The main purpose for using this stratification variable was to ensure sufficient sampling precision for each of the states. This variable, therefore, was used as an explicit stratification variable. Consequently the frame was divided into eight separate lists, one for each state.

2. **School type** – This variable refers to the three major education sectors in Country X: Government, Catholic, and Independent schools. In past studies, this variable provided substantial gains in sampling precision and was therefore used as an implicit stratification variable. Implicit stratification was also an efficient means of ensuring adequate representation of all three school types in the sample.   Within each explicit stratum, defined by state, the schools were grouped by education sector.

Country X could have considered urbanization (rural, urban) as a second implicit stratification variable. This would have produced possibly 6 (3x2) implicit strata within each explicit stratum. In this case, the schools within each education sector would be additionally split into rural and urban groups.

All of this information should be reported in *Sampling Form 3*.

---

5. STEP 3 — NATIONAL SAMPLING PLAN

5.1. Introduction

The development of national sampling plans is the responsibility of National Project Managers. The national sampling plans will need to be adapted to get the particular requirements of national education systems. The staff at AIR will be an invaluable resource to the NPMs in developing their national sampling plans. NPMs will need to get their national sampling plans approved by AIR, prior to the selection of the school samples

AIR has set high standards consistent with other international studies for sampling precision, participation rates, and implementation of sampling plans. These standards will result in samples of the highest quality and subsequent survey estimates that are unbiased, accurate, and internationally comparable.

> ► All samples must be sent to AIR for approval by **May 15, 2001** and must be approved by AIR by **June 1, 2001**.

5.2. Effective Sample Size

The AIR standard for sampling precision requires that all national samples achieve an effective sample size, for the main criterion variables, of at least 400 school sites (for an infinite population).

The number of school sites was selected so as to meet a sampling precision requirement equivalent to that of other well-known international studies.

An effective sample size of 400 school sites results in the following approximate 95% confidence limits for sample estimates of population means, percentages, and correlation coefficients:

- *Means*: m +/- 0.1 s (where "m" is a school mean estimate and "s" is its estimated standard deviation);

- *Percentages*: p +/- 5% (where "p" is a school-level percentage estimate); and

- *Correlations*: r +/- 0.1 (where "r" is an estimated school correlation).

AMERICAN
INSTITUTES
FOR RESEARCH

5.3. Computing the Actual Sample Size: Adjustments Using the Finite Population Correction (FPC)

Although the precision requirements of the study require an effective sample size of 400 schools, in many countries the actual sample size will differ because of the total number of schools in the country. It is well known that for a random sample of size n from an infinite population, the variance of the mean is σ^2/n. When the population is finite (i.e., when we know the total number of units in the population) the factor $(N-n)/N$ is introduced in computing the variance. This factor is called the finite population correction and its effect is to adjust the variance based on the known population size.

When applied to the determination of sample size, the finite population correction has the effect of reducing the number of sampled units needed to achieve a given precision level. Appendix B presents a table listing the actual number of schools to be sampled in a country given its total number of schools. Countries should look up the number that corresponds most closely to the number of schools in their country and find the corresponding number of schools to be sampled. This number of sample school is the minimal number of schools a country is expected to sample. Countries can sample additional units if they wish to increase the precision of their estimates, especially in the use of explicit strata.

> ► NOTE: NPMs must look up the actual number of schools to be sampled based on the total number of schools in their country in the Sample Size Table provided in Appendix B.

5.4. Participation Rates

Consistent with other international studies, AIR requires a minimum participation rate of 50% of initially sampled schools and 85% after replacement schools are included. These standards exist to minimize the potential for nonresponse biases.

The use of replacement schools does not guarantee that potential response biases have been reduced. NPMs are encouraged to persuade as many sampled schools as possible to participate in the Main Survey. The selection and use of replacement schools are described in Sections 2.5, 7.5, and 7.9.2.

AMERICAN
INSTITUTES
FOR RESEARCH

► Only a high participation rate will minimize the potential for nonresponse biases. For tips on how to increase participation, refer to Section 4 of the *USSS Survey Operations Manual* Ref. No. 01-01-01.

5.5. Reporting Participation Rates

National participation rates will be presented in the AIR analytical reports. They will include weighted and unweighted school participation rates with and without replacement schools.

The AIR analytical results will be annotated based on each national center's ability to meet the stated participation rate targets. National samples that meet these requirements only after including replacement schools will be annotated in the international reports. National samples that do not meet these requirements, even after including replacement schools, will be so indicated.

5.6. Sample Design Framework

The general sample design framework adopted for this study is a one-stage stratified sample design. National sampling plans must rely on sound and defensible sampling methods. Schools can be stratified, explicitly and/or implicitly, and are selected with probabilities proportional to their size (PPS). The sampling method is called a PPS systematic sampling method and is described in Chapter 7.

5.7. Small Schools

The presence of small schools tends to be problematic when developing a sampling plan. A small school is one where the total in-program enrollment in the school is smaller than the average in-program school enrollment in the country (ASEC) divided by 4 (ASEC4). Any school where enrollment is less than ASEC4 is considered a small school.

Problems arise with small schools in computing sampling weights. Since their probability of selection is very small, they tend to have very large weights that may lead to large fluctuations (that is, large standard errors) in the indices based on the sample data. The more small schools there are, the greater this problem becomes.

NPMs who can demonstrate that enrollment in small schools represents less than 10% of all national enrollment in upper secondary schools need not worry about the presence of small

schools. Small schools should remain on the school sampling frame and be subject to normal sampling procedures.

If, however, enrollment in small schools exceeds 10%, NPMs should define an explicit stratum of small schools.

NPMs may choose to exclude small schools (MOS < ASEC4) if the resulting exclusion rate is less than 2%. This is described in *Sampling Form 4*. NPMs should consider this alternative only if their overall exclusion rate remains below the 5% limit set by AIR.

NPMs should consult AIR to resolve any potential issues arising from the presence of small schools and the required information should be reported in *Sampling Form 4*. Based on this information, AIR will determine whether a particular national sampling plan will require special treatment of small schools.

AMERICAN
INSTITUTES
FOR RESEARCH

6. STEP 4 — THE SCHOOL SAMPLING FRAME

6.1. Introduction

This step requires NPMs to construct a school sampling frame for their national defined target population. A well-constructed school sampling frame is one that provides complete coverage of the nationally defined target population without being contaminated by incorrect entries, duplicate entries, or entries that refer to elements that are not part of the defined target population.

The construction of a school sampling frame depends to a great extent on the availability of appropriate information about school sites. The main types of sampling frames that are likely to be used in this study are described below. NPMs will be asked to describe their choice of a sampling frame in *Sampling Form 5*.

6.2. Single-Level Sampling Frame

Single-level sampling frames are based on a comprehensive national list of school sites. A measure of size (MOS) is associated with each school listed in the sampling frame. The MOS gives a fairly accurate indication of the number of students in each listed school. The preferred MOS would be the number of students currently enrolled in the target programs. Current enrollment data, however, is rarely available at the time schools are sampled. The quality of a sampling frame will, to a large extent, depend on the accuracy of the MOS available. The MOS should be recorded on *Sampling Form 6*, the *School Tracking Form*.

A single-level sampling frame consists of separate lists of schools for each explicit stratum. Each school is represented by an entry on one, and only one, of those lists. Each school entry should include at a minimum:

- A unique school identification information, such as a national school ID, name, or address;

- Appropriate entries for all explicit and implicit variables; and

- A suitable school MOS.

In addition, because many questions will address program issues, NPMs should make every effort to obtain school enrollment at the program level.

Each list can be sorted according to a suitable set of implicit stratification variables, if implicit stratification is to be used.

6.3.   Double-Level Sampling Frame

Double-level sampling frames are used when a comprehensive list of national school sites is not available. As a consequence, double-level sampling frames introduce an additional stage of sampling before the sampling of school sites. Although this manual does not explicitly provide instructions to implement this additional stage of sampling, the school sampling methods presented in this manual can be adapted for that purpose. AIR will provide support for NPMs that require the use of a double-level sampling frame.

The use of a double-level sampling frame increases the degree of clustering in the sample. This could be cost effective in countries where distances are great and/or travel costs are high. Double-level sampling frames, however, usually require a larger sample size because of this increased clustering. NPMs are encouraged to make every effort to locate a comprehensive national list of schools that could produce a single-level sampling frame.

6.3.1. First Level

The first level of this type of sampling frame usually consists of reliable information on student enrollment associated with geographical areas (GAs). The GAs are usually identifiable geographical units used during the taking of a national census, or administrative units for which basic educational statistics are available.

The GAs become the primary sampling units (PSUs) at this first sampling stage. There needs to be a sufficiently large number of PSUs to permit the random selection of a minimum number of PSUs in the sample. This minimum number will ensure that the sampled PSUs can effectively be considered representative of all PSUs. The AIR guidelines require a minimum of 100 PSUs in the first sampling frame and a minimum of 50 PSUs sampled from that frame. NPMs will need to discuss the implementation of these guidelines with AIR, based on their specific circumstances. The inability to meet these guidelines will mean that the sampled PSUs can only be considered representative of themselves, rather than of all PSUs. The result would be a potentially low national coverage, which would be annotated in the analytical reports.

The first level sampling frame consists of a list of all eligible PSUs. Each PSU in the list should include at minimum:

- PSU identification number;

- Appropriate entries for any suitable implicit stratification variables; and

- A suitable PSU measure of size.

The list can be sorted according to a suitable set of implicit stratification variables.

A suitable PSU measure of size could be:

- Student enrollment in the target programs;

- Total student enrollment;

- Number of schools; or

- Population size.

6.3.2. Second Level

The second level of this type of sampling frame consists of comprehensive lists of schools from all sampled PSUs. For each PSU selected from the first level sampling frame, a comprehensive list of schools is prepared according to the general format used for a single-level sampling frame.

6.4. Other Types of Sampling Frames

The use of other types of sampling frames is not recommended. AIR plans to produce school-level results, so other types of sampling frames may not be suitable for this. NPMs who wish to use a sampling frame that is not related to either types described above must consult AIR to determine the suitability of their sampling frame.

6.5. School Measure of Size

A suitable school measure of size (MOS) is a critical component of a school sampling frame, since school selection probabilities are based on this MOS. Possible school MOS, in decreasing order of suitability, are:

1. Total student enrollment in the target programs in a given school;
2. Total student enrollment in the target school;
3. Average student enrollment per grade, i.e., total student enrollment divided by the number of grades in the school; and
4. Number of classrooms in the target school.

The enrollment data should be as current as possible. If none of these types of MOS are available, or if the available enrollment data is too out-dated, schools may have to be selected with equal probabilities. The systematic sampling method can still be used.

► NPMs must describe the type of sampling frame they intend to use.

If a double-level sampling frame, or any other type, is to be used, NPMs must provide a description of the information available to construct this frame. For a double-level sampling frame, this information should describe the PSUs and their total number, the PSU MOS available, and the number of PSUs to be sampled.

NPMs should also describe the school MOS they intend to use.

6.6.   Excluding Schools from the Sampling Frame

Based on the information reported in *Sampling Form 2, National Defined Target Population*, some schools may need to be excluded from the school sampling frame. These schools must be identified in *Sampling Form 7*, along with the reason for their exclusion.

NPMs should keep in mind that the proportion of eligible students in all excluded schools must not exceed 5% of the national desired target population. NPMs may have chosen to exclude extremely small schools. This choice may well depend on being able to keep exclusions within the 5% limit. Excluding extremely small schools is not recommended if this will cause the exclusion rate to rise above 5%.

6.7.   Sorting the Sampling Frame

In Step 2 (Chapter 4), NPMs defined a set of implicit stratification variables as well as a sequence from which they will be used to form implicit strata. Through this process, each school in the sampling frame will be assigned a value for each implicit stratification variable. For example, if urbanization (rural or urban) and school type (public or private) are used, each school must be classified as rural or urban and as either public or private.

The school sampling frame is then sorted by the implicit stratification variables. The schools are first sorted by the first implicit stratification variable, then by the second implicit stratification variable, and so on, until all implicit stratification variables have been exhausted. The result is a cross-classification structure the represents the implicitly stratified school sampling frame. The maximum number of implicit strata resulting from this process is equal to the product of the

number of different levels of each implicit stratification variable. The realized number of implicit strata may be less than the maximum. This is because some cross-classifications could containing no schools at all. For example, there may not be any rural private schools.

As an example, using urbanization and school type, the cross-classification will yield a maximum of four cells (2x2): rural-public, rural-private, urban-public, and urban-private. If private schools are only found in urban centers, then the rural-private cell will be empty.

Within each implicit stratum, the schools must be sorted by their MOS. The sort order should alternate from high to low and then low to high, as indicated in the following table:

Implicit Stratum	Sort Order of MOS
1. Rural-Public	High to Low
2. Rural-Private	Low to High
3. Urban-Public	High to Low
4. Urban-Private	Low to High

For more than four implicit strata, the sort order pattern in the table should simply be continued.

6.8.   Treatment of Small Schools

In Step 3 (Chapter 5), NPMs specified the extent to which small schools are represented in their sampling frame. If the presence of small schools becomes important, the sampling frame may require special treatment.

Information on the prevalence of small schools must be reported in *Sampling Form 4*. This form also provides some guidelines about the treatment of small schools. NPMs should consider three guidelines in the treatment of small schools in their country:

1. If the proportion of eligible students enrolled in small schools (MOS < ASEC4) is less than 2%, then NPMs may choose to exclude them from their sampling frame. This will increase the size of their excluded population, so NPMs must be sure that they still meet the target of no more than 5% of exclusion from all sources.

2. If the proportion of eligible students enrolled in small schools is less than 10%, then NPMs need not take any further action regarding their small schools. Small schools should remain on their sampling frame and NPMs can proceed with the usual school sampling activities.

3. If the proportion of eligible students enrolled in small schools is greater than 10%, then NPMs need to form an explicit stratum of small schools since they will need to be treated differently from the other schools in the sampling frame. Through certain tabulations, it could be determined that small schools are of concern only for a subset of the sampling frame, for example among rural schools or private schools. This would only require, for example, the formation of an explicit stratum of small schools in the rural stratum. The MOS of all small schools would be set equal to the average MOS in that stratum, resulting in an equal probability sample of small schools. The allocation of schools to the explicit strata should remain proportional to student enrollment. This will avoid fluctuations in the sampling weights. All sampled small schools should be included.

6.9. Sampling Frame Statistics

NPMs are required to document the process of sampling frame construction. In addition to the descriptive information reported in *Sampling Forms 3, 4,* and *5*, NPMs should record basic statistics by strata in *Sampling Form 8*. These statistics should be tabulated from the school sampling frame. For each combination of explicit and implicit strata, the number of schools and enrolled students should be reported. These statistics will be useful for finalizing the sample allocation of schools to explicit strata and as control totals to validate the computation of sampling weights.

► NPMs must list and describe all combinations of explicit and implicit strata in *Sampling Form 8*. For each entry, they should tabulate the number of schools and eligible students. These statistics should be readily available from the school sampling frame.

6.10. School Sample Allocation

Based on the discussion of school sample sizes presented in Chapter 5 and the information reported in *Sampling Form 8*, NPMs should now be able to allocate their sample of schools to the explicit strata. This sample allocation should be reported in *Sampling Form 9*.

NPMs are responsible for allocating their school sample. NPMs should consult AIR for any assistance in completing this task.

NPMs should pay particular attention to the following points in allocating their national school sample:

- If an explicit stratum of small schools is required, NPMs will need to allocate their school sample accordingly.

- NPMs must ensure that at least two schools are selected for the Main Survey in each explicit stratum.

6.11. Reminder: Sampling Timeline

All school samples, both for the Field Test and the Main Survey, must be selected and approved by their respective deadlines, **March 15, 2001** and **June 1, 2001**.

For the sake of efficiency, it is important that NPMs submit all required sampling forms on time. AIR will respond as quickly as possible to NPMs once these forms have been received and reviewed.

► NPMs who are concerned they might not meet the deadlines for school sample selection should notify AIR as soon as possible.

---

7.   STEP 5 — THE SAMPLE OF SCHOOLS

7.1. Overview

This chapter describes how to select the sample of schools. It is presented as a series of operational steps leading to the selection and identification of all sampled schools. The school sampling method is described as a PPS systematic sampling method.

The school selection probabilities are based on the school measure of size (MOS). This sampling method can be applied to both single-level sampling frames as well as double-level sampling frames. With double-level sampling frames, the second level is like a single-level sampling frame. This sampling method can easily be adapted to the first level of a double-level sampling frame.

For each sampled school, two candidates are identified as replacement schools. Although the replacement schools serve as possible substitutes for non-cooperating schools, NPMs should keep their use to a minimum by ensuring the highest possible participation rate for sampled schools.

Schools are selected once the school sampling frames have been compiled for each explicit stratum. Each school entry in the sampling frame should contain:

- A school identification number;

- Values for any implicit stratification variables; and

- A measure of size.

The school sampling frame should be sorted according to any implicit stratification variables. School samples will be selected separately for each explicit stratum.

Although this manual provides paper versions of the sampling forms for recording all sampling results, NPMs are expected to rely on computer software to perform the sample selection and to generate the required sampling forms. NPMs need not send the exact sampling forms provided in this manual. They may generate their own versions, either on paper or as computer files, as long as the forms or computer files include all of the information shown on the sampling forms in this manual.

AMERICAN
INSTITUTES
FOR RESEARCH

7.2. Determining Which Schools to Sample

Sampling Form 10 is used to identify which schools are selected in the sampling frame. It is used to perform the calculations that will enable the NPMs to select the sample of schools for the Main Survey. This form generates a series of selection numbers that will be applied to the sampling frame to identify the sampled schools in *Sampling Form 11*.

7.2.1. Calculating the Sampling Interval

The PPS systematic sampling method requires the computation of a sampling interval for each explicit stratum. This calculation is based on information recorded in *Sampling Form 11*. Calculations will be done separately for each explicit stratum. Be sure to indicate the explicit stratum being considered in each case.

To calculate the sampling interval:

- Record the total measure of size "S" for all schools in the sampling frame for the specified explicit stratum. This number is obtained from column (4) of *Sampling Form 8*. Record this number in box [a] of *Sampling Form 10*.

- Record the number of schools "D" to be sampled from the specified explicit stratum. Record this number in box [b] of *Sampling Form 10*.

   Note: Enter sequentially the numbers from 1 through "D" in column 1 (line numbers) of *Sampling Form 10*.

- Calculate the sampling interval "I" as follows: $I = S \div D$

- Record the sampling interval "I" to four decimal places in box [c] of *Sampling Form 10*.

7.2.2. Generating a Random Start

Select a random number for each explicit stratum. Random numbers are now easily generated by computer software. NPMs may also rely on random number tables found in some sampling books. NPMs who generate these random numbers using software should be sure to use a "uniform random number generator," which generates numbers distributed uniformly between 0 and 1.

The generated random number should be recorded in box [d] of *Sampling Form 10*. The selected random number should be between 0 and 1 and recorded to four decimal places.

AMERICAN
INSTITUTES
FOR RESEARCH

7.2.3. Calculating the Selection Numbers

Selection numbers must be calculated and recorded in column (2) of *Sampling Form 10* for each line number specified, i.e., 1 through "D."

- To obtain the first selection number, multiply the sampling interval "I" by the random number "RN." Record this first selection number to four decimal places in column (2) of the first line, corresponding to line number 1, in *Sampling Form 10*. This first selection number will be used to identify the first sampled school in the specified explicit stratum.

- To obtain the second selection number, simply add the sampling interval "I" to the first selection number. The second selection number will be used to identify the second sampled school. Record this number to four decimal places in column (2) of the second line.

- Continue adding the sampling interval "I" to the previous selection number to obtain the next selection number. This is done until all specified line numbers (1 through "D") have been assigned a selection number.

In summary, the first selection number in an explicit stratum is RN x I; the second selection number is (RN x I) + I; the third selection number is (RN x I) + I + I; and so on.

Selection numbers need to be generated independently for each explicit stratum. Be sure to select new random numbers for each explicit stratum. For each explicit stratum, the first line number in column (1) of *Sampling Form 10* should always be 1 (one).

► NPMs must complete *Sampling Form 10* to demonstrate how the school sample is selected.
   Separate sampling forms are required for each explicit stratum.

7.3. Identifying the Sampled Schools

Based on the selection numbers recorded in *Sampling Form 10*, the sampled schools may now be identified on the school sampling frame. Since selection numbers were selected separately for each explicit stratum, it is important to match the correct selection numbers to the explicit strata in the sampling frame.

The task, however, is to compile a cumulative MOS on the sampling frame. This cumulative MOS will determine which schools are sampled and is tabulated within explicit strata. This means the cumulative MOS must be reset when moving from one explicit stratum to the next.

All of the relevant school sampling frame information is recorded in *Sampling Form 11*. It is expected that the contents of *Sampling Form 11* will exist in the form of a computer file, perhaps Excel, dBase, SAS, or SPSS. All of these formats are acceptable and NPMs are encouraged to send such files to AIR as a record of their sample selection activities rather than paper versions of *Sampling Form 11*.

Sampled schools are identified as follows:

> Let "Z" denote the first selection number for a particular explicit stratum. Find the first school in the sampling frame whose cumulative MOS equals or exceeds "Z." This will be the first sampled school. In other words, if C' is the cumulative MOS of a particular school in the sampling frame and C" is the cumulative MOS of the school immediately preceding it, then the schools whose cumulative MOS is C' is sampled if:
>
> - C' is greater than or equal to Z, and
> - C" is strictly less than Z.
>
> Apply the above rule to all selection numbers for a given explicit stratum and enter a check mark (√) for each sampled school in column (6) of *Sampling Form 11*. The line number corresponding to the selection number used should also be recorded in column (5).

7.4. Very Large Schools

NPMs may discover that their sampling frame contains schools whose MOS is greater than the calculated sampling interval. Conceptually, this would mean that some of these very large schools are sampled schools more than once. This situation can be problematic because it complicates the data base design and computation of sampling weights.

In order to avoid these problems, it is recommended that NPMs construct an explicit stratum of very large schools. This stratum would contain all schools whose MOS is greater than the calculated sampling interval. This will likely require a number of iterations, where a new sampling interval needs to be computed at each iteration. Since this new sampling interval will be smaller than the previous one, there may be more schools whose MOS is greater than the newly calculated sampling interval. In practice, two or three iterations will suffice to complete the construction of the explicit stratum of very large schools.

All schools in the explicit stratum of very large schools will be selected for the Main Survey.

AMERICAN
INSTITUTES
FOR RESEARCH

> ► If NPMs notice schools in their sampling frames whose MOS is greater than the computed
> sampling interval, they will encounter sampling difficulties since such schools could be
> selected more than once. If this occurs, NPMs should notify AIR for advice on how to
> resolve this issue.

7.5. Identifying Replacement Schools

It is not always possible to obtain the participation of all sampled schools. In order to avoid the
resulting sample size losses, a mechanism to identify replacement schools for non-participating
sampled schools is needed. It is important to identify replacement schools at the same time as
the sample schools are drawn to avoid the haphazard use of alternate schools as replacements,
which may amplify response biases. Although this approach does not necessarily avoid non-
response bias, it tends to reduce the potential for bias. Furthermore, it is conceptually more
palatable than over-sampling to accommodate a low participation rate.

NPMs must make every effort to get as many of the sampled schools to participate as possible.
In some cases, however, NPMs will need to consider the use of replacement schools. To allow
for this, each sampled school in the Main Survey will be assigned two replacement schools in the
sampling frame.

Replacement schools for the Main Survey are identified as follows. For each sampled school,
identified by a check mark (√) in column (6) of *Sampling Form 11*, the next two schools
immediately following it in *Sampling Form 11* are designated as its replacement schools. They
should be labeled respectively "R1" and "R2" in column 6 of *Sampling Form 11*. The use of
implicit stratification variables, and the subsequent ordering of the school sampling frame by
size, ensures that any sampled school's replacements will have similar characteristics.

Should a sampled school happen to be the last school listed in an explicit stratum, then the two
schools immediately above it are its designated replacement schools. Should a sampled school
happen to be the next to last school listed in an explicit stratum, then the schools immediately
above it and immediately below it are its designated replacement schools.

> ► A sampled school cannot, under any circumstance, be designated as the replacement school
> for another sampled school. Should the above described procedures produce a sampled
> school that is also designated as a replacement school, the NPM should contact AIR for
> advice on how to resolve the problem.

NPMs may find it difficult to assign replacement schools to some large sampled schools because the sampled schools occur very close to each other in the sampling frame. In this situation, NPMS may consider schools above sampled schools as replacement schools. At times, NPMs may only be able to assign a single replacement school, or even none, when two consecutive schools in the sampling frame are sampled. NPMs should make every effort to get such large sampled schools to participate so as not to rely on replacement schools.

7.6. Ineligible Schools

An ineligible school is a sampled school that is not eligible to participate in the survey. This can occur when a school has closed since the sampling frame was created. This can also occur when a sampled school does not have any students in the target program, such as an elementary school erroneously included in the sampling frame. The discovery of ineligible schools among sampled schools should be rather unusual. It is important to keep in mind that ineligible schools must not be replaced. Replacing ineligible schools will likely lead to an over-estimation of the national effective target population.

The presence of ineligible schools in a national sample could significantly reduce the overall sample size if a good number of them are found. NPMs should make every effort to remove all ineligible schools from the national school sampling frame, and to get the most up-to-date sampling frame, prior to sampling. NPMs who suspect the presence of a non-trivial amount of ineligible schools in their sample should consider increasing their school sample size to compensate for the potential loss of sampled schools.

► If a sample school is ineligible, it must not be replaced.

 If a first replacement school is ineligible, it must be replaced by the designated second replacement school.

7.7. Assigning School IDs

In order to keep track of sampled schools and replacement schools in the NPM's and the AIR databases, all sampled schools and replacement schools must be assigned unique identification numbers. The school IDs must be assigned to all sampled schools and replacement schools and recorded in column (7) of *Sampling Form 11* and in column (1) of the *School Tracking Form* (refer to the next section of this manual and to the *USSS Survey Operations Manual* Ref. No. 01-01-01 for further information about the *School Tracking Form*).

School IDs should be assigned according to the following numbering procedure:

- You should number the schools in the designed sample sequentially, beginning with 0001. For example, if 300 schools are sampled for the Main Survey, then they are assigned IDs from 0001 through 0300.

- For the replacement schools in the first replacement sample (R1) of the Main Survey, you should add 1000 to the School IDs of the sample schools they would replace. Similarly, for the replacement schools from the second replacement sample (R2), you should add 2000 to the School IDs of the schools in the sample. For example, the first replacement school for the school with School ID 0014 would have School ID 1014 and the second replacement school would have School ID 2014.

► NPMs must complete *Sampling Form 11* as a written record of their sampling frame and school sample selection. *Sampling Form 11* should preferably be submitted as a computer file with all relevant information included. There must be as many *Sampling Form 11*s as there are explicit strata.

NPMs should be particularly careful when assigning schools IDs since this will be the only method of recognizing sampled schools and their corresponding replacement schools.

7.8. Creating a School Tracking Form

To monitor the participation status of sampled schools, all sampled schools and their replacement schools should be recorded in *Sampling Form 6*, the *School Tracking Form*. This form will contain all contact and scheduling information, along with basic sampling information. Replacement schools assigned to a sampled school should also be recorded in this Sampling Form. There is one form for each sampled school and its replacement schools.

NPMs can start the preparation of *Sampling Form 6* once the sample of schools is selected. This form, however, will only be completed once data collection has ended since only then will NPMs know which sampled schools participated and which ones required replacement. *School Tracking Form*s for Main Survey schools should be submitted as soon as possible after the data collection for the Main Survey.

► NPMs must complete *Sampling Form 6* to keep track of the participation status of all sampled schools and their replacement schools. NPMs can use this form to record contact information.

AMERICAN
INSTITUTES
FOR RESEARCH

7.9.   Confirming School Participation

NPMs must confirm and document the participation status of sampled schools. This is done using *Sampling Form 6*. In column (1), the USSS school ID is recorded. The corresponding school MOS as reported in the sampling frame , is recorded in column (2). Any contact information available is recorded in columns (3) and (4).

7.9.1. Sampled Schools

NPMs must make every effort to obtain the participation of as many sampled schools as possible. This is important in order to minimize the potential for non-response biases and to meet the AIR requirements for school participation rates. Please refer to Chapter 4 of the *USSS Survey Operations Manual* Ref. No. 01-01-01 for procedures that will help to ensure the highest participation rate possible. Sampled schools should be listed in the first row of *Sampling Form 6*.

Once a sampled school's participation has been obtained, a check mark (√) should be entered in column (5) of the first row. If a sampled school is unwilling to participate then enter "N" in column (5).

7.9.2. Replacement Schools

NPMs will need to contact replacement schools for each of the sampled schools that will not participate (after all efforts to obtain participation of the sampled school has failed). All sampled schools that will not participate should be replaced, unless they are ineligible schools (see Section 7.6). If a first replacement school does not participate, or is ineligible, then NPMs must use the designated second replacement school.

7.10.   Computing Response Rates

As mentioned earlier, the response rate requirements for this study are that a minimum of 50% of originally selected schools and a minimum of 85% of total schools after replacement participate in the survey to avoid flagging of country results in the reports. NPMs are expected to use *Sampling Form 12* to report the response rates in their country.

Computing the response rate involves 2 steps:

1. ensuring that at least 50% of original schools participate; and
2. ensuring that at least 85% of schools participate, after replacements.

To document step 1, enter the total number of originally sampled schools in box [a] of *Sampling Form 12*, followed by the number of originally sampled schools that agreed to participate in box [b]. The original school participation rate is computed simply by dividing the number in box [b] by the number in box [a], and reporting it in box [c].

For step 2, the study response rate, again enter the total number of originally sampled schools in box [d] of *Sampling Form 12*. This time, in box [e], enter the total number of schools that agreed to participate, including replacement schools. The study response rate is computed simply by dividing the number in box [e] by the number in box [d], and reporting it in box [f].

8. STEP 6 — COMPUTING AND CHECKING THE SAMPLING WEIGHTS

8.1. Introduction

Appropriate estimation of population characteristics must take into account any disproportionate representation in the sample due to the sample selection process. This is accomplished by assigning a weight to each school, where the weights approximately account for the sample design and reflect the appropriate proportional representation of the various types of units in the population. The computation of sampling weights involves the following steps:

1. Computing the school probability of selection;

2. Adjusting this probability for non-response; and

3. Computing the final school weight.

The school probability of selection is simply the probability that a sampled school was included in the sample. Non-response adjustment adjusts this probability to account for those selected schools that refused to participate. Finally, the final school weight is the reciprocal of the adjusted probability of selection.

► NOTE: Sampling weights must be attached to each data record (i.e., each responding school).

The sampling weights should be calculated separately for each school. The steps for calculating these weights are provided in the following sections. NPMs must use *Sampling Form 13* to document the weighting process. In *Sampling Form 13*, the sampled schools should be listed in the order in which they were selected. Also, record the school identification number in column (1) and whether the school participated in the study in column (2). If a replacement school was used, record the information of the replacement school, not the originally selected school. All computations should be recorded to four decimal places.

8.2. Computing the School Probability of Selection

The school probability of selection is simply the probability of selecting the school for the sample. For each school in the sample, the probability of selection should be computed as:

AMERICAN
INSTITUTES
FOR RESEARCH

$$P_i = \frac{mAi}{A}$$

where m is the number of units in the sample, Ai is the measure of size (MOS) for school i, and A is the aggregate measure of size (AMOS) for the population within the explicit stratum, that is, the sum of the size measures for all schools in the frame that are in the explicit stratum containing the sampled school.

The school measure of size, the aggregate measure of size for the population, and the school probability of selection should be recorded in *Sampling Form 13* in columns (3), (4), and (5), respectively.

Example 3   **Probability of Selection**

To illustrate the calculation of these probabilities, consider a sample of 300 schools in a country where the total enrollment in in-scope upper secondary programs is 260,384 and where there is no explicit stratification. School 1 has a measure of size of 576. Hence the school probability of selection is:

$$P_i = \frac{300(576)}{260,384} = 0.6636$$

For School 2, the measure of size is 95. Thus, for this school, the probability of selection is:

$$P_i = \frac{300(95)}{260,384} = 0.1095$$

► NOTE: The computation of the selection probabilities, adjusted probabilities, and school weights must be carried out separately for each explicit stratum.

8.3.   Adjusting the Probabilities of Selection for Non-Response

To compensate for nonparticipating schools, the school probabilities of selection should be adjusted in the following way:

1.   In box [a] of *Sampling Form 13*, enter L, the number of schools selected for sampling.

---

2. In box [b] of *Sampling Form 13*, enter L', the number of schools that participated by returning a filled out questionnaire.
3. Calculate the adjustment factor by dividing L'/L and rounding out the result to two decimal places. Enter the rounded value in box [c] of *Sampling Form 13*.
4. For each school listed in *Sampling Form 13*, multiply the school probability of selection in column (5) by the "adjustment factor" in box [c]. These values are the "adjusted probabilities of selection," and will be denoted by the symbol *H* in column (6).

Example 4   **Adjusted Probability**

To illustrate the calculation of the adjusted school probabilities of selection, let's assume again that we have our sample of 300 schools. Out of these 300 schools, 261 of them participated in the study. One would enter L=300 in box [a] of *Sampling Form 13*, and L'=262 in box [b]. The adjustment factor would simply be 262/300=0.8733. You would report it in box [c].

Then to use our two schools from Example 3, their adjusted probability of selection would simply be:

   For school 1: (0.6636)(0.8733)=0.5795

   For school 2: (0.1095)(0.8733)=0.0956

8.4. Calculating the School Weights

The school weight is simply the inverse of the adjusted probability of selection. Thus, for each school listed in *Sampling Form 13*, compute the school weight by:

$$W_i = \frac{1}{H_i}$$

where *Hi* is simply the adjusted probability of selection from column 6. Report this final school weight (*Wi*) in column (7) of *Sampling Form 13*.

Example 5   **Final School Weights**

Again, using our earlier example, the final school weights for the 2 schools would be:

   For school 1: 1/0.5795=1.7256

   For school 2: 1/0.0956=10.4603

8.5. Checking the Computation of the School Weights

The sum of the school weight * measure of size for each school in the sample should be approximately equal to the aggregate measure of size for the population, which is reported in column (4). Hence, you can verify the accuracy of your computations by checking this number against the aggregate measure of size. The sum of the school weight * measure of size for each school in the sample should be reported in box [d] of *Sampling Form 13*. This can easily be computed using the following formula:

$$\sum w_i * MOS_i$$

where w_i is the weight for each sampled school and MOS_i is the measure of size for each sampled school.

► NOTE: NPMs are especially encouraged to submit the data for *Sampling Form 13* as an electronic file.

AMERICAN
INSTITUTES
FOR RESEARCH

Appendix A:

Sampling Forms

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 1
National Desired Target Population

See Section 3.3 of the USSS School Sampling Manual.

Country:

National Project Manager:

1. Total national enrollment in upper secondary programs listed in
    Phase I Classification Study:

    [a]

2. Describe the population(s) to be excluded from the national desired target population (if
    applicable):

3. Total enrollment excluded from the national desired target population
    (corresponding to the exclusions listed in the previous item):

    [b]

4. Total enrollment in the national desired target population:
    box [a] – box [b]

    [c]

5. Percentage of coverage in the national desired target population:
    box [c] ÷ box [a]

    [d]

6. Describe your data source *(provide copies of relevant tables)*:

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 2
National Defined Target Population

See Section 3.4 of the USSS School Sampling Manual.

Country:

National Project Manager:

1. Total national enrollment in the national desired target population:
    From box [c] on Sampling Form 1

    [a]

2. School-level exclusions:

Description of exclusions	# of students
TOTAL	[b]

Percentage of school-level exclusions:
box [b] ÷ box [a]

%

3. Total enrollment in the national defined target population:
    box [a] – box [b]

    [c]

4. Describe your data source *(provide copies of relevant tables)*:

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 3
Stratification

See Section 4.4 of the USSS School Sampling Manual.

Country:

National Project Manager:

Explicit Stratification

1. List and describe the variables used for explicit stratification.

	Explicit stratification variables	# of levels
1		
2		
3		
4		
5		

2. Total number of explicit strata:

Implicit Stratification

3. List and describe the variables used for implicit stratification.

	Implicit stratification variables	# of levels
1		
2		
3		
4		
5		

4. Total number of implicit strata:

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 4
Small Schools

See Section 5.7 of the USSS School Sampling Manual.

Country: _____

National Project Manager: _____

Total enrollment _____
Total # of schools _____
ASEC _____
ASEC4 _____

1.　　　Enrollment in small schools:

Type of school based on enrollment	# of students	Percentage
In-program school enrollment < ASEC4		[a]
In-program school enrollment ≥ ASEC4 and < ASEC		[b]
In-program school enrollment ≥ ASEC		[c]
TOTAL………………………………………………..		100 %

2.　　　If the percentage in box [a] is less than 2%, then these small schools can be excluded from the national defined target population. Be sure to record this exclusion on *Sampling Form 2*, item 2.

　　　　box [e] < 2%?　　　O Yes　　　*or*　　　O No

3.　　　If the percentage in box [c] is 90% or more, then all schools should remain in the national defined target population and be subject to normal school sampling.

　　　　Box [c] ≥ 90%?　　　O Yes　　　*or*　　　O No

4.　　　If the percentage in box [c] is less than 90%, then small schools (enrollment < ASEC4) should form an explicit stratum. Be sure to record this level of explicit stratification in *Sampling Form 3*, Item 1. Please consult with AIR to determine an appropriate sample for this stratum of small schools.

MAIL COPIES OF ALL FORMS TO:
Dr. Eugene Johnson, The American Institutes for Research, Pelavin Research Center
1000 Thomas Jefferson Street, NW, Washington, DC, 20009-3835, USA

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 5
Sampling Frame Description

See Chapter 6 of the USSS School Sampling Manual.

Country: _____

National Project Manager: _____

1. Specify the type of school sampling frame to be used.
   O Single-level sampling frame
   O Double-level sampling frame
   O Other *(please describe)*: _____

   _____

   _____

   _____

2. Specify the school measure of size (MOS) to be used.
   O Total student enrollment in the target programs in the given school.
   O Total student enrollment in the target school
   O Average student enrollment per grade
   O Number of classrooms in the target school
   O Other *(please describe)*: _____

   _____

   _____

   _____

3. Specify the school year for which enrollment data will be used for the school MOS.

   _____

   _____

   _____

4. If a double-level sampling frame is to be used, please provide a preliminary description of the information available to construct this frame. Please consult with AIR for support and advice in the construction and use of a double-level sampling frame.

   _____

   _____

   _____

   _____

   _____

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 7
Excluded Schools

Use additional sheets as necessary. See Section 6.6 of the USSS School Sampling Manual.

Country:

National Project Manager:

(1) School ID	(2) Reason for exclusion	(3) School MOS

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 8
Population Counts by Strata

Use additional sheets if necessary. See Section 6.9 of the USSS School Sampling Manual.

Country:

National Project Manager:

(1) Explicit Strata	(2) Implicit Strata	(3) Population Counts Schools	(4) Students

MAIL COPIES OF ALL FORMS TO:

Dr. Eugene Johnson, The American Institutes for Research, Pelavin Research Center
1000 Thomas Jefferson Street, NW, Washington, DC, 20009-3835, USA

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 9
Sample Allocation by Explicit Strata

Use additional sheets as necessary. See Section 6.10 of the USSS School Sampling Manual.

Country:

National Project Manager:

(1) Explicit Strata	(2) Sample Allocation Schools	(3) Students

MAIL COPIES OF ALL FORMS TO:

Dr. Eugene Johnson, The American Institutes for Research, Pelavin Research Center
1000 Thomas Jefferson Street, NW, Washington, DC, 20009-3835, USA

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 10
School Sample Selection

Use additional sheets as necessary. See Section 7.2 of the USSS School Sampling Manual.

Country:

National Project Manager:

Explicit Stratum:

S	D	I	RN
[a] Total Measure of Size	[b] Desired Sample Size	[c] Sampling Interval	[d] Random Number

(1) Line Numbers	(2) Selection Numbers

Page___of___

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 11
School Sampling Frame

Use additional sheets as necessary. See Section 7.3 of the USSS School Sampling Manual.

Country:

National Project Manager:

Explicit Stratum:

(1) School List ID	(2) Implicit Stratum	(3) MOS (measure of size)	(4) Cumulative MOS	(5) Line Number	(6) Sample Status	(7) USSS School ID

Page___of___

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 6
School Tracking Form

Use one form for each sampled school. See Sections 6.2, 7.8, and 7.9 of the USSS School Sampling Manual.

Country: _____

National Project Manager: _____

Explicit Stratum: _____

	(1) USSS School ID	(2) School MOS	(3) Name, Address, and Telephone # of School	(4) Name and Telephone # of School Coordinator	(5) Status*	(6) Date Materials Sent	(7) Date Materials Received
Sampled School							
First Replacement School							
Second Replacement School							

* Enter "N" for non-participating schools. A check mark (√) indicates the participating school.

Page ____ of ____

MAIL COPIES OF ALL FORMS TO:

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 13
Calculating the School Weights

Country: _____

National Project Manager: _____

Explicit Stratum: _____

L	L'					
[a] Number of selected schools	[b] Number of responding schools	[c] Adjustment factor (L'/L)			[d] Sum of weights ($\sum w_i$)	

(1) School ID	(2) Participated (Y/N)	(3) MOS	(4) AMOS	(5) School probability, P_i	(6) Adjusted school probability, H_i	(7) Final school Weight, W_i

Use additional sheets as necessary.

Page ____ of ____

OECD International Survey of Schools at the Upper Secondary Level
Sampling Form 12
Response Rate

See Section 7.10 of the USSS School Sampling Manual.

Country: _____

National Project Manager: _____

1.   Originally selected schools

[a] Total selected schools	[b] Original participating schools	[c] School response rate [c] = [b] ÷ [a]

2.   Total sample

[d] Total selected schools	[e] Total participating schools including replacements	[f] Study response rate [f] = [e] ÷ [d]

Appendix B:

Sample Size Table

OECD International Survey of Schools at the Upper Secondary Level
Sampling Size Table

Number of schools in country	Number of schools needed for sample
100	80
150	109
200	134
250	154
300	172
350	187
400	200
450	212
500	222
550	232
600	240
650	248
700	255
750	261
800	267
850	272
900	277
950	282
1000	286
1100	294
1200	300
1300	306
1400	311
1500	316
1600	320
1700	324
1800	327
1900	331
2000	333
3000	353
4000	364
5000	370
6000+	400