

SCHOOL SAMPLING PREPARATION MANUAL

PISA 2012 MAIN SURVEY

Doc: SchSamp_MS12_2

Consortium:

Australian Council for Educational Research (ACER, Australia)

cApStAn Linguistic Quality Control (Belgium)

Deutsches Institut für Internationale Pädagogische Forschung (DIPF, Germany)

Educational Testing Service (ETS, USA)

Institutt for Lærerutdanning og Skoleutvikling (ILS, Norway)

Leibniz - Institute for Science and Mathematics Education (IPN, Germany)

National Institute for Educational Policy Research (NIER, Japan)

The Tao Initiative: CRP - Henri Tudor and Université de Luxembourg - EMACS
(Luxembourg)

Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe, Belgium)

Westat (USA)

TABLE OF CONTENTS

1. Introduction.....	3
1.1 A Note about Changes from PISA 2009.....	3
1.2 A Note about Versions	3
1.3 A Note about Appendices	3
1.4 Relationship of Field Trial Sampling Procedures to Main Survey Sampling Procedures.....	4
1.5 The Responsibilities of National Project Managers and the Consortium in Sampling Schools.	5
2. Summary of NPM School Sampling Main Tasks	7
3. Timeline	8
4. Acronyms	9
5. Main Survey Sampling and PISA Standards.....	10
5.1 Target Population and Sampling	10
5.2 Participation Rates	11
5.2.1 Schools.....	12
5.2.2 Students.....	14
5.2.3 Reporting Participation Rates.....	15
6. Overview of PISA Main Survey Sample Design.....	15
7. Main Survey Sampling Tasks.....	19
7.1 Sampling Task Seven: Submitting Your Population Size and Exclusions	19
7.1.1 Sampling Task Seven A (ST7A): National Desired Target Population	19
7.1.2 Sampling Task Seven B (ST7B): National Defined Target Population	21
7.1.2.1 School-Level Exclusions.....	22
7.1.2.2 Within-School Exclusions.....	23
7.2 Sampling Task Eight: Submitting Your Sampling Frame and School-Level Exclusions	27
7.2.1 Sampling Task Eight A (ST8A): Description of School Sampling Frame	27
7.2.2 Sampling Task Eight B (ST8B): Sampling Frame and School Exclusion	28
7.2.2.1 School-Level Sampling Frame	28
7.2.2.2 Area-Level Sampling Frame	31
7.2.2.3 School-Level Exclusions.....	32
7.3 Sampling Task Nine: Finalize Sample Size Requirements.....	32
7.4 Sampling Task Ten: MS School Sample Delivery	33
7.5 Sampling Task Eleven: Reviewing and Agreeing to SFKQ.....	33
7.6 Main Survey Student Sampling Tasks.....	36
7.6.1 Obtaining Student Lists from Participating Schools.....	36
7.6.2 Learning the KeyQuest Functions and Operations Required for Student Sampling and Performing Student Sampling in KeyQuest	38
7.7 Sampling Task Twelve: Submitting School Participation Information and Validity Checks to Westat	39
Appendix A - Treatment of Small Schools	43
Appendix B - School Sampling Selection Methodology	47
B.1 Strata Numbering.....	47
B.2 School Sample Allocation Over Explicit Strata	47
B.3 Assigning a Measure of Size (MOS) to each School	48
B.4 Sampling Frame Statistics	48
B.5 Sorting the Sampling Frame	49
B.6 Determining Which Schools to Sample.....	50
B.6.1 Calculating the Sampling Interval.....	50

<i>B.6.2 Generating a Random Start</i>	51
<i>B.6.3 Calculating the Selection Numbers</i>	51
B.7 Certainty School Stratum	52
B.8 Identifying the Sampled Schools	52
B.9 Identifying Replacement Schools	53
B.10 Assigning STIDSTRT and STIDSCH	54
Appendix C – Examples of Completed Sampling Forms	56
Appendix D - Sampling Form Templates	76
Appendix E – Financial Literacy and Computer-based Assessments	91
E.1 Financial Literacy	91
E.2 Computer-based Assessments	93
E.2.1 Student Sample Size	93
E.2.2 School Sample Size	94
E.2.3 Non-Participation.....	94
E.2.4 SEN Students.....	100
Appendix F – The Link to PISA 2003 International Option	101
F.1 Overview	101
F.2 The Overlap Control Method	101
Appendix G – Participation in Grade or other National Option Sampling and Sample Size	105
G.1 Samples of Students in a Particular Grade.....	105
G.2 Other National Options	109
Appendix H – Area-level Sampling Frames	110
Appendix I - What is a School?	112
Appendix J - Within-School Exclusion Weighted Estimates for PISA 2009 Countries	113

1. Introduction

1. This document provides guidelines for the submission and approval of sampling and population information required prior to and shortly after the PISA 2012 Main Survey (MS). Note that the structure of this manual is similar to the PISA 2012 Field Trial (FT) Guidelines.

1.1 A Note about Changes from PISA 2009

2. Most aspects of this manual are similar to those of the PISA 2009 manual. Where a change of consequence has been made from the 2009 manual, this is noted in the margin as “New for 2012”. Major differences between the FT and the MS, e.g., the treatment of small schools, are also highlighted in the margins.

1.2 A Note about Versions

3. Before the PISA MS is completed, there may be two or more versions of this manual produced. The expected difference between these versions is that later versions will include additional material to help inform NPMs about the sampling-related activities of the Consortium and to deal with special topics of interest to only some NPMs.
4. In cases where there is a correction or other important change, from one version to the next, NPMs will be explicitly notified about this. We do not expect that NPMs should have to search through each version to find out if there are important differences that affect them.
5. However, please be sure to replace the manual with a new version when it arrives, and **be sure that you are using the latest version of the manual at all times.**

1.3 A Note about Appendices

6. This sampling preparation manual generally limits non-Appendix sections to what needs to be done, how it needs to be done, and when it needs to be done, in addition to some overview material.
7. Information about the PISA 2012 adapted treatment of small schools can be found in Appendix A. This should be reviewed since small schools affect most countries
8. Since the Consortium selects the MS school sample, details about school sample selection can be found in Appendix B. Particular areas in Appendix B that should be reviewed are sections B.9 and B.10 so that replacement schools can be better understood.
9. Appendices C and D contain examples of completed sampling task forms and sampling form templates respectively.

New for 2012

10. Countries participating in any of Computer-Based assessments, or the Financial Literacy Option, can read in Appendix E about how the Consortium will deal with these for MS school sampling. The few countries participating in the Reading Components option will have an appendix on this topic sent to them just after the Sampling Preparation Manual is released.
11. Countries indicating to the Consortium that they desire maximal overlap with PISA 2003 can read about how the Consortium will accomplish this in Appendix F.
12. Countries selecting a grade-based sample or a sample of other national option non-PISA students can find discussion about how this will be accomplished in Appendix G.
13. Aside from these, additional appendices are: H, about area-level sampling frames; I, regarding how we try to define a school; and J, that contains PISA 2009 weighted estimates of within-school exclusion categories. These estimates in Appendix J should only be used for estimates of sampling form ST7B within-school exclusions if there are no available national data that are more recent.

1.4 Relationship of Field Trial Sampling Procedures to Main Survey Sampling Procedures

14. It is very important to note the similarities and differences between the sampling procedures for the FT and the MS, because while some aspects are very similar between the two, others are very different. Simply following the FT procedures on a larger scale for the MS will, in most cases, result in a MS sample that is completely unacceptable.
15. The procedures for listing and sampling students within schools will be very similar between the FT and MS. The only changes will be enhancements that are developed in the procedures as a result of the FT, or the introduction of some national requirements that did not exist for the field trial.
16. The reasons for the differences between the FT and the MS are that the two components have very different purposes. Recall that the FT had two main purposes:
 1. To collect data to ensure that the instruments developed for the MS contain test and questionnaire items that are sound in all countries and that they are properly translated; and
 2. To test the operational procedures for sampling students and conducting assessments within schools.
17. The MS, on the other hand, has the purpose of obtaining a data file that will permit analyses that give valid statistical inferences about the PISA student population, and the characteristics of the schools that they are in. Countries will obtain data that will be comparable with all other countries participating in PISA. To provide valid and comparable estimates of student

achievement and characteristics, the sample of students needs to be selected using established and professionally recognised principles of scientific sampling, in a way that represents the full population of PISA-eligible students. These purposes impose more rigorous requirements on the sampling procedures than those used in the FT. In particular, the development of a complete and up-to-date list of schools, and sampling from this list, requires much more care and preparation for the MS than was necessary for the FT. In addition, a larger sample of schools is required at the MS stage.

1.5 The Responsibilities of National Project Managers and the Consortium in Sampling Schools

18. NPMs are responsible for the following tasks. Some of these have already been done at the FT stage and are noted with an “*”.

1. Establishing the age definition, based on birth date, and the time of testing, according to PISA rules, so as to ensure that the correct student population is surveyed, and to ensure that the assessment can be completed in the designated time period.*
2. Determining the school, and student level exclusions that will apply in their country in PISA, ensuring that these exclusions are kept to a minimum, and documenting the nature and quantity of these exclusions.
3. Obtaining, and where necessary enhancing, a list of schools and other educational institutions in the country that will contain the population of enrolled students that are to be covered by PISA.
4. Identifying **at least three months before the time of MS school sample delivery**, any national options, including any possible oversampling, not previously discussed that could have an impact on sampling.
5. Determining the sample sizes of schools and students that are needed, in conjunction with Westat and ACER, to satisfy international PISA requirements and any additional national requirements.
6. Identifying suitable stratification variables for the schools on the sampling frame*, and ensuring that these variables are present and correct for all schools.
7. Submitting the school sampling frame to MyPISA so that Westat and ACER can select the school sample.
8. Maintaining accurate records as to which sampled schools participate in PISA, the reason for each school that does not participate, and the use of replacement schools in the sample where appropriate.

New for 2012

19. The Consortium, and Westat and ACER in particular, are responsible for the following tasks:

1. Checking, from the FT stage, that each country has identified an appropriate age definition and time of testing.
2. Checking that the exclusions in each country are clearly defined, necessary, and minimal.

3. Assisting each country in determining the sample size and stratification that will meet both PISA and national objectives.
 4. Selecting the school sample and returning it to the NPM, with documentation.
 5. Developing school and student sampling and replicate weights to permit valid inferences to be made from the sample to the population.
 6. Developing measures of school and student response and exclusion, and coverage of the target population, to assist in evaluating the comparability and quality of the achievement and other data from each country.
20. For some aspects the manual may document, usually in an appendix, the reason for certain procedures and requirements. This is not the primary purpose of the manual, and NPMs may need to consult other PISA documents (such as minutes of various meetings) if they wish to understand in detail the purpose of a particular requirement. The prime purpose of the manual is to provide instructions for completing the necessary school sampling tasks.
21. As noted, this sampling preparation manual provides the mechanism by which the Consortium informs NPMs what they need to do to carry out their responsibilities. The sampling forms that are included with the manual provide paper version templates to indicate the type of information that will be collected. However, all the information will be collected via MyPISA as for the FT. Westat and ACER will use the submitted MyPISA information to carry out their responsibilities for monitoring and quality assurance, school sample selection, and later weighting. Note that although ACER will select the school samples, **all school sampling enquiries should be directed to Westat.**
22. As will be explained when discussing each form in section 7, some of these forms need to be prepared by the NPM and sent to the Consortium via MyPISA for review, while others will be first prepared by the Consortium and sent to the NPM for review.
23. The data from these forms will be used in assessing the quality of the PISA data, and therefore need to be as accurate as possible.
24. Following the completion of discussion about the FT sampling forms, a proposed schedule for submitting the MS sampling forms, linked to the proposed MS testing dates, was negotiated with each national centre.
25. If any changes are required to this schedule, please inform Westat as soon as possible to assist with resource planning.

2. Summary of NPM School Sampling Main Tasks

26. NPMs are expected to complete the following main tasks:

- Meet with Westat personnel at the NPM meeting as needed or as requested;
- Attend any available KeyQuest training session at the NPM and coder training meetings especially if a country or NPM is new and/or chooses to participate in optional components;
- Identify early to Westat any not yet discussed possible oversampling or other national options that could affect sampling;
- Submit MS sampling information on population size, and exclusions;
- Submit excluded schools and sampling frame;
- Finalize sample size requirements with Westat and check school frame processing;
- Receive and confirm receipt of the MS school sample, and check the sample;
- Review and approve the SFKQ;
- Collect student lists;
- Review the KeyQuest functions and operations required for student sampling; and implement student sampling;
- Submit school participation information and validity checks to Westat.

3. Timeline

Note: The numbering of sampling tasks for the MS starts from Sampling Task Seven, as a continuation of the sampling tasks from the FT. This is done because some FT forms collected information for both the FT and the MS and may also be referenced during the MS.

Oct 9-Oct 14, 2011 – At the NPM meeting in Singapore you should meet with Westat and ACER sampling personnel as recommended or desired. You should also attend any available KeyQuest training especially if you are a new country or NPM and/or if you choose to participate in any optional components.

Due three months before the negotiated MS school sample delivery date, you should let Westat know, via email, of any planned oversampling or other national options that could impact sampling, and that have not already been discussed with Westat or noted on the Optional and Variations on-line form on MyPISA (Administration Tasks). If your national option is not finalized by the time the MS school sample should be selected, **the MS school sample will be selected for PISA to meet the negotiated delivery date**. If the national option is later finalized and more sampling needs to be done by the Consortium, a delivery date will be negotiated, and additional sampling costs may be incurred.

New for
2012

Due two months before the negotiated MS school sample delivery date you should complete the following sampling tasks:

Sampling Task Seven A (ST7A) - you should submit MS sampling information on the initial (desired) target population;

Sampling Task Seven B (ST7B) - you should submit MS sampling information on the final (defined) target population;

Sampling Task Eight A (ST8A) - you should submit the description of the MS sampling frame;

Sampling Task Eight B (ST8B) - you should submit the MS sampling frame and excluded schools.

By one month before the negotiated MS school sample delivery date (Sampling Task Nine (ST9)): you should receive your sample allocation and Small Schools Analysis from Westat. Set aside adequate time to check frame processing and to finalize sample size requirements with Westat.

At the negotiated MS school sample delivery date (Sampling Task Ten (ST10)), you should receive your MS school sample from Westat. You should confirm your receipt of the sample and should also check the sample to ensure you have a clear understanding of the various files you received. You should ask any questions that you have about the sample at this time.

Due one month after the negotiated MS school sample delivery date (Sampling Task Eleven (ST11)), you should review and agree to the ST11 which will become SFKQ. This form must be agreed prior to student sampling¹.

About eight weeks prior to the first testing date, you should prepare for and collect the student listings. You should also review the KeyQuest function and operations required for student sampling; and implement student sampling in KeyQuest once the lists are received.

Due one month after the end of the data collection period (Sampling Task Twelve (ST12)), you should submit school participation information and validity checks to Westat - **reminder or not**.

¹ In addition, the demographic tables must have been agreed, and your manuals must have been verified. Refer to DMM.

4. Acronyms

CBA	Computer-based assessment
CBAL	Computer-based Assessment of Literacies option
CMOS	Conditional MOS used in overlap control procedures
CPROB	Conditional probability used in overlap control procedures
DMM	Data Management Manual
E3COM	A variable on the <i>School Participation</i> KeyQuest instrument to explain irregular school status situations
ENR	Approximate enrolment of PISA eligible students in each school
ERA	Electronic Reading Assessment
FL	Financial Literacy option
FLUH	Financial Literacy UH (FLUH) booklet option for SEN students
FT	Field Trial
GAs	Geographical Areas used as first-stage sampling units
HLM	Hierarchical Linear Modeling
MAS	Manuals Adaptation Spreadsheet
MOS	Measure of Size used in PPS sampling
MS	Main Survey
PBA	Paper-based assessment
PGB	PISA Governing Board
PPS	Probability Proportional to Size
PROBP	PISA probability of Selection
PROBI	Probability of selection for the other survey involved in overlap control
PS	Computer-based Problem Solving component
PSUs	Primary Sampling Unit
RD	Indicator for dual role replacement schools on the school frame
SEN	Special Education Needs student
SCP	<i>School Participation</i> instrument in KeyQuest
SFKQ	Created from ST11, the sampling form for KeyQuest
SSA	Small School Analysis
STF	Student Tracking Form
ST1	Cyclewide Sampling Task 1
ST2	Cyclewide Sampling Task 2
ST7A	MS Sampling Task 7A; information on the Initial (desired) Target Population
ST7B	MS Sampling Task 7B; information on the Final (defined) Target Population
ST8A	MS Sampling Task 8A; information about the sampling frame
ST8B	MS Sampling Task 8B; the sampling frame and school exclusions
ST9	MS Sampling Task 9; finalizing sample size and allocation
ST10	MS Sampling Task 10; sample selection by ACER and Westat
ST11	MS Sampling Task 11; reviewing the form that will become SFKQ
ST12	MS Sampling Task 12; submitting school participation and validity checks to Westat
TCS	Target Cluster Size
UH	The “Une-Heure” (One-Hour) paper-based booklet and short questionnaire option for SEN students

5. Main Survey Sampling and PISA Standards

5.1 Target Population and Sampling

27. A number of definitions and standards are listed in the Technical Standards for PISA 2012. Meeting the standards specified in this section will ensure that the students tested come from a comparable target population in every country, and are in an equivalent age range. Therefore, the results obtained will not be confounded by potential age effects. Furthermore, to be able to draw conclusions that are valid for the entire population of PISA students, a representative sample needs to be selected for participation in the test. The size of this representative sample should not be too small, in order to achieve a certain precision of measurement in all countries. For this reason, minimum numbers of participating students and schools are specified.

Standard 1.1 The PISA Desired Target Population is agreed upon through negotiation between the National Project Manager and the International Contractor, within the constraints imposed by the definition of the PISA Target Population.

Standard 1.2 Unless otherwise agreed upon only PISA-Eligible students participate in the test.

Standard 1.3 Unless otherwise agreed upon, the testing period:

- is no longer than six consecutive weeks in duration,
- does not coincide with the first six weeks of the academic year, and
- begins exactly three years from the beginning of the testing period in the previous PISA cycle.

Standard 1.4 Schools are sampled using agreed upon, established and professionally recognised principles of scientific sampling.

Standard 1.5 Students are sampled using agreed upon, established and professionally recognised principles of scientific sampling and in a way that represents the full population of PISA-Eligible students.

Standard 1.6 The PISA Defined Target Population covers 95% or more of the PISA Desired Target Population. That is, school-level exclusions and within-school exclusions combined do not exceed 5%.

Standard 1.7 The student sample size is a minimum of 4,500 assessed students for PISA participants and 1,500 assessed students for additional adjudicated entities, or the entire PISA Defined Target Population where the PISA Defined Target Population is below 4,500 or 1,500 respectively.

Standard 1.8 The school sample size is a minimum of 150 schools for PISA participants, and 50 schools for additional adjudicated entities, or all schools that have students in the PISA Defined Target Population where the number of schools with students in the PISA Defined Target Population is below 150 or 50 respectively.

Standard 1.9 The school response rate is at least 85% of sampled schools. If a response rate is below 85% then an acceptable response rate can still be achieved through agreed upon use of replacement schools.

Standard 1.10 The student response rate is at least 80% of all sampled students across responding schools.

Note 1.1 The Target Population and Sampling standard apply to the Main Survey but not the Field Trial.

Note 1.2 Data from schools where the student response rate is greater than 25 % will be included in the PISA dataset.

Note 1.3 For the purpose of calculating school response rates, a participating school is defined as a sampled school in which more than 50 % of sampled students respond.

Note 1.4 Guidelines for acceptable exclusions that do not affect standard adherence, are as follows:

- School level exclusions that are exclusions due to geographical inaccessibility, extremely small school size, administration of PISA would be not feasible within the school, and other agreed upon reasons and that total to less than 0.5 % of the PISA Desired Target Population,
- School level exclusions that are due to a school containing only students that would be within-school exclusions and that total to less than 2.0 % of the PISA Desired Target Population, and
- Within-school exclusions that total to less than 2.5 % of the PISA Desired Target Population– these exclusions could include, for example, students not able to do the test because of a functional disability.

Note 1.5 Principles of scientific sampling include, but are not limited to:

- The identification of appropriate stratification variables to reduce sampling variance and facilitate the computation of non-response adjustments.
- The incorporation of a target cluster size of 35 PISA-Eligible students which upon agreement can be increased, or reduced to a number not less than 20.

28. Work towards satisfying standards 1.1, 1.2, and 1.3 was started at the FT stage and will continue through the MS stage. Standards 1.4, 1.7 and 1.8 are the main topics of this manual. Standard 1.5 is one of the main topics of the Data Management Manual (DMM) and also briefly discussed in this manual. Standard 1.6 will be discussed in a later section of this manual. Standards 1.9 and 1.10 are discussed next.

5.2 Participation Rates

29. The PISA data quality standards require minimum participation rates for schools, as well as for students. These standards exist to minimise the potential for response biases.

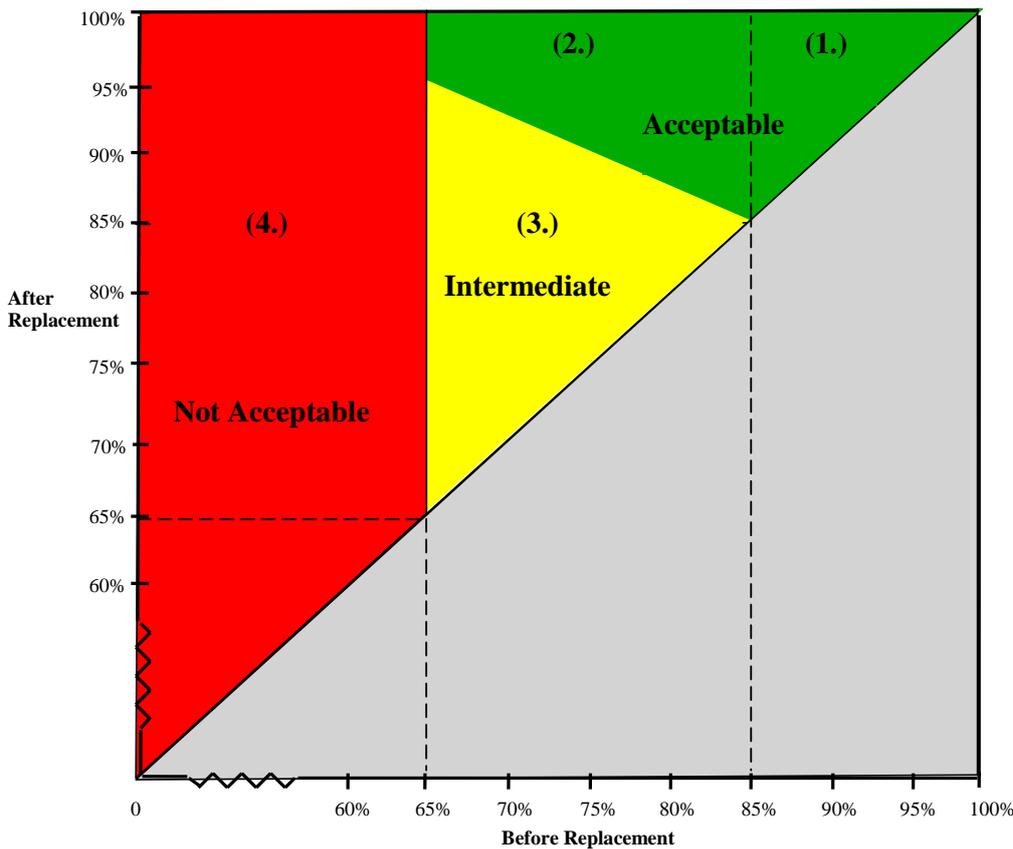
Different
from the FT

5.2.1 Schools

30. PISA requires a minimum weighted participation rate of 85% of originally sampled schools. However, nonparticipating sampled schools may be substituted with “replacement schools” to meet sample size and response rate requirements. The use of replacement schools does not guarantee that potential biases have been reduced. Therefore, NPMs are encouraged to persuade as many original sampled schools as possible to participate; only a high participation rate will minimize the potential for response bias. The identification and use of replacement schools are described in Appendices B.9 and B.10, and in Table 3, section 7.7. Note that raising participation rates, also known as response rates, through the use of replacements improves quality, but that a given response rate achieved through the use of replacements is not as good as that same rate achieved without replacements. That is, a school participation rate of x% before replacement has less potential for biases than the same response rate of x% achieved only after replacement. Therefore, all other factors being equal, acceptability of the country's data in international comparisons will be relative to both the school participation rate of originally sampled schools and the response rate achieved with the use of replacements.
31. Three response rate zones -- acceptable, intermediate and not acceptable -- are defined. “Acceptable” means that the country’s data will be included in international comparisons. “Not Acceptable” means that the country’s data will be a candidate for not being reported in international comparisons, and will be included only if the NPM provides considerable evidence that nonresponse bias is likely to be minor. Finally, the “Intermediate” zone means that a decision on whether or not to include the country’s data in international comparisons will be made while taking into account various other factors. These zones are a function of the response rate before replacement, and the response rate after replacement. There are three sets of combinations of before and after response rates that are noted as being in the “acceptable” zone.
32. Consider some examples.
- A country with a before replacement response rate of 87% would have its data considered “acceptable” according to this criterion, even before any school level replacements were used.
 - A country with a before replacement response rate of 60% would have its data “not acceptable” no matter what the after replacement response rate was.
 - A country with a before replacement response rate of 70%, and an after replacement response rate of 96% would have its data considered “acceptable” according to these criteria.
 - A country with a before replacement response rate of 75% and an after replacement response rate of 91% would have its data considered “acceptable” according to these criteria.

- A country with a before replacement response rate of 70%, and an after replacement response rate of 90% falls into the intermediate zone. Such a set of conditions requires a decision on whether or not to accept the data from the country. Other factors that will be considered in such a situation include student-level response rates, exclusion rates, quality control data, and closeness of the response rates to the acceptable level.
- The response rate zones are shown diagrammatically in the following graph. Note that the “acceptable” region has part of its region bounded by the Before Replacement Rate $> 85\%$ vertical line, and another part consisting of the After Replacement Rate $> 95\%$ horizontal line for Before Replacement Rates between 65% and 85%. The final part of the acceptable region is defined by the Before Replacement Rate between 65% and 85% and After Replacement Rate $> (255\% - \text{Before Replacement Rate})/2$.
- Note that a school with less than 50% participation among the selected eligible and non-excluded students will not be considered as a participating school. If such a school has less than 25% participation among the selected eligible and non-excluded students, then the students in such schools will not be included in analysis. If such a school has more than 25% and less than 50% participation among the selected eligible students, the students in these schools will be retained for analysis, even though the school is considered a non-participant for the purposes of monitoring response rates.
- Note that “selected eligible students” are those in the PISA target population, and therefore participation measures are relative to only those students in the target population. For example, participation will be determined separately for the PISA-eligible portion and the grade-eligible portion of a national grade option sample in the same school. This could result in the school deemed as participating (50% or more of selected grade-eligible students participated) for the grade sample, but as a nonparticipating school for PISA if less than 50% of the selected PISA-eligible non-excluded students participate.
- Also note that a selected PISA-eligible student recorded in the database as not doing the cognitive part of the PISA assessment will be counted as a nonparticipant for the 50% and 25% participation measures, even if the student completed the background questions of the student questionnaire.

PISA 2012 School Response Rates



5.2.2 *Students*

33. PISA also requires a minimum participation rate of 80 percent of students within participating schools (sampled and replacement). This applies in aggregate, not to each individual school. Follow-up sessions may be required in schools where too few students participated in the originally scheduled test sessions so as to ensure a high overall student response rate. Guidelines for determining when follow-up sessions are appropriate are found in the PISA NPM Manual for the 2012 MS located under [Materials on MyPISA](#). NPMs are responsible for determining whether follow-up sessions are required, but may delegate this task to the School Coordinators and Test Administrators. As they oversee the actual test sessions, they will be able to promptly determine whether follow-up sessions are required at each school.

34. Student participation rates are calculated over all participating schools, whether sampled schools or replacement schools, and from the participation of students at the originally scheduled session and any follow-up session that may be required. The student participation rate requirement needs to be met at the national level, not necessarily for each participating school. Student response rates will be calculated separately for PISA paper-based and CBA, if CBA was a country option

35. Replacement students are not utilized in PISA.

5.2.3 *Reporting Participation Rates*

36. National participation rates will be presented in the PISA analytical reports. They include:
- Weighted and unweighted school participation rates with and without replacement schools (see 5.2.1 for a discussion of the minimum weighted response rates required); and
 - Weighted and unweighted student participation rates (minimum 80% required, weighted).
37. The PISA analytical results for each country will be annotated, based on whether or not the response rate requirements have been adequately met.

6. Overview of PISA Main Survey Sample Design

38. The overall sampling plan for PISA 2012 for the MS is similar to that used in previous PISA cycles. That sampling plan begins with the definition of the PISA Target Population.
- The Target Population for PISA starts with students attending **ALL** educational institutions located within the country, and in grade 7 or higher. The “standard” PISA target population is further refined to its age basis: students between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the testing period.
 - Aside from the “standard” population definition, and because of the allowed one month variation, the population could be slightly younger: between 15 years and two (completed) months and 16 years and 1 (completed) months at the beginning of the chosen testing period, or it could be slightly older: between 15 years and 4 (completed) months and 16 years and 3 (completed) months at the beginning of the chosen testing period.
 - When the PISA Target Population is henceforth referenced, it needs to be understood that the population could be defined by any one of the possible birth date definitions for the specified testing period.
39. The international desired target population (i.e., the PISA Target Population) is intended to provide full coverage of all PISA-eligible students (defined by the birth date definition) in a country’s education system. This means that countries need to include (as well as any PISA-eligible students attending regular programmes) PISA-eligible students who attend school on a part-time basis, are in vocational training or other non-general types of programmes, or any other related type of educational programme, or who are in foreign schools within the country, **even if they are not included in other international or national studies**. The desired target population does not include residents who attend school in a foreign country. **Therefore, all schools located within a country with the potential to have PISA-eligible students in grades 7 or above at the time of assessment need to be made available for sampling from a complete listing of such schools.**

40. The population of PISA schools is limited to schools that could contain PISA-eligible students at the time of the assessment. Schools that could not contain any such students, such as lower primary schools in some countries, for example, are therefore not considered part of the school-level population.
41. As a reminder from the FT, recall that since the largest part of the PISA target population is made up of 15-year-olds, then “15-year-olds” is the term often used when referring to the PISA target population **even though the PISA population actually includes one to three months of age 16.**
42. Since PISA is a survey of students, and as the sampled students have to be given an assessment under standard conditions, in every PISA country it is clear that the sample of students should be selected by first selecting a sample of schools, and then selecting a sample of students within those schools. Thus the need for careful, rigorous, standardised, and documented sampling applies both to the selection of schools, and students within schools.
43. The sample design proposed for the PISA assessment is generally referred to as a two-stage stratified design. The first-stage sampling units consist of individual schools having any possibility of having PISA-eligible students at the time of assessment. Schools are sampled systematically with probabilities proportional to a measure of size (PPS), with the measure of size being a function of the estimated number of PISA-eligible students enrolled (ENR). Sampling with PPS provides a practical technique when sampling from populations with large variation in the values of the study variables, and often gives a desirable gain in efficiency. The efficiency of PPS sampling depends on the availability of an auxiliary measure which is expected to be strongly related to the student performance. For PISA, this auxiliary measure is the ENR variable.
44. PPS sampling of schools followed by equal probability sampling of students within sampled schools has desirable design qualities. This design results in student weights that are approximately equal (within each stratum). Low variability in student weights leads to more precise sample estimates compared to highly variable student weights.
45. The comprehensive national list of all eligible schools is called the school sampling frame. Prior to sampling, schools in the sampling frame can be assigned to a predetermined number of explicit strata (mutually exclusive groups of schools which together cover the whole school sampling frame) and/or implicit strata (variables for sorting schools in the explicit strata, or the whole school frame if no explicit strata are used). If explicit strata are used, the school sample is allocated over the explicit strata in proportion to the PISA students in each stratum. A minimum of 150 schools will be selected in each country, with the requirements of national options often requiring a somewhat larger sample. Note that the minimum of 150 schools is

selected with the expectation that there will be at least 150 **participating** schools, once field exclusions, ineligibility and nonresponse are accounted for. As the schools are sampled, replacement schools are simultaneously identified, should they be needed to replace eligible, non-excluded and non-participating sampled schools.

46. More than 150 schools usually need to be selected because of small schools. Small schools are a concern because fewer than the desired number of students is sampled from such schools. This leads to a reduction in student sample size. Therefore, small schools may require special treatment as well as an overall increase in the school sample size. This issue is discussed in more detail in Appendix A.
47. Countries may need to sample more than the minimum number of required schools, to meet national requirements. For example, in countries with highly tracked school systems at this age (where track is not used in stratification), the school sample needs to be appropriately increased. This is to allow for the considerable school-to-school variation in achievement that can be anticipated. Such an increase is not required, however, in highly tracked school systems at this age, if the different tracks are included in distinct explicit or implicit strata. For this reason, stratification by school level and/or track is strongly encouraged.
48. Countries may also wish to increase their samples for regional or other comparisons. As this oversampling is one form of a national option, sample sizes for regions or other school types need to be discussed and negotiated with Westat at least three months before the MS school sample is to be delivered. When oversampling for regional or other comparisons, it is even more important that the school ENR values for schools in each oversampled region or school type are as accurate as possible. If there is doubt about the accuracy of the ENR variable, then the school sample size should be increased from the initial negotiated size to compensate for the uncertainty in the ENR measures.
49. Some small countries may not have 150 schools that could contain PISA students. In such cases, a census of the PISA schools will be taken. Care must be taken to ensure that all schools that could have PISA students are included on the school list when it is submitted to MyPISA. **Schools later discovered cannot be added to the sample.** This applies also to strata in any country where a school census will be conducted.
50. The second-stage sampling units are students within sampled schools, and are to be sampled using the Consortium software, KeyQuest. Once schools are selected to be in the sample, a list of each sampled school's PISA-eligible students needs to be prepared. A within-school sample size, the 'target cluster size' (TCS), is nominated by each PISA participant through Sampling Task 2 (ST2). The TCS is the number of students that are to be sampled from schools with large PISA enrolment. This number of eligible students is sampled with equal

New for 2012

probability from each of the schools sampled for PISA², assuming there is at least this number of eligible students at the school. **In sampled schools where there are fewer eligible students than the TCS, all of these students are sampled. An assessment needs to be done at such schools, even if there are only a few PISA students.** The usual TCS is 35 students. In total, a minimum sample size of 4500 assessed students is to be **achieved**. Note that 30 students (150 schools x 30 students = 4500 students) has been inflated to 35 to account for an anticipated combined nonresponse, ineligibility, and exclusion rate of 15% among students within schools. It is possible for countries to negotiate a different TCS as a national option, but if it is reduced from 35 then the sample size of schools has to correspondingly be increased beyond 150, so as to ensure that at least 4500 students in total will be sampled and assessed.

51. The TCS has to be at least 20 so as to ensure adequate accuracy in estimating variance components within and between schools—an important analytical objective of PISA.

Different
from the FT

Schools with fewer than 20 PISA students must be included on the frame and if sampled, need to have their students assessed.

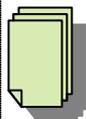
52. Note that the TCS and the actual PISA student sample size are not necessarily the same thing. All schools in the country usually have the same TCS value. At the time when school lists of PISA students are collected, some schools thought to be large may have fewer than TCS PISA students. In such cases, the actual PISA student sample size will include all PISA students, and this will be smaller than the TCS.

53. For countries that participated in previous cycles of PISA and which had larger than anticipated sampling variances associated with their estimates, or had fewer than 4500 assessed students, recommendations will be made about sample design changes that will help to rectify these problems for PISA 2012. Additionally, although information was supplied on FT sampling forms, new countries will be asked through email exchanges about anticipated school and student response rates, as well as about rates of both school and student ineligibility based on their FT experience. Westat can take this information into account when determining school sample sizes to guard against low student yield.

² For different explicit strata, it is possible to vary the target cluster size.

7. Main Survey Sampling Tasks

7.1 Sampling Task Seven: Submitting Your Population Size and Exclusions



Due two months before the negotiated MS school sample delivery date, or as agreed

54. This task involves two subtasks. Subtask A requires you to submit MS sampling information on your initial national desired target population using the form for Sampling Task 7A (MS Initial Target Population). Subtask B requires you to submit the information on your final national defined target population using the form for Sampling Task 7B (MS Final Target Population). **You will need to submit this information by uploading documents to MyPISA for approval by Westat.**

7.1.1 *Sampling Task Seven A (ST7A): National Desired Target Population*

55. Every NPM needs to define and describe their country's national desired target population (or initial population). The national desired target population definition should address the requirements of the **international** desired target population. In fact, the PISA Governing Board (PGB) and the Consortium strongly encourages all NPMs to provide complete national coverage of their international desired target population.

56. In cases where the national desired target population deviates from full national coverage of all PISA eligible students, these deviations should be described and enrolment data provided to measure the extent of the reduced coverage. Such situations may occur if NPMs find it necessary to reduce their national coverage by removing, for instance, a small, remote geographical region, or a language group, possibly due to political, organizational or operational reasons. Areas deemed by the PGB to be part of a country (for the purpose of PISA) but which are not included for sampling will be designated as non-covered areas, and need to be documented.

57. If such situations occur, the national desired target population will differ from the international desired target population. If a significant proportion of students are excluded from the full national population of PISA-eligible students, this will mean that survey results will not be deemed representative of the entire national school system. Therefore, the Consortium will work with NPMs to try to keep such differences minimized, before these types of exclusions are approved.

58. Mention needs to be made about how to treat home schooled students, distance education students, and open enrolment students. In any of these cases, if the students are not enrolled in a physical school, then they are ineligible for PISA because they do not attend a physical educational institution where they could be assessed if sampled. If such students are not included in the PISA population, a comment about these students needs to be added to the comment box of the ST7A form. If such students sometimes attend an institution, or attend at least for national exams, they are part of the PISA population.
59. **The initial national desired target population information should be described on the Sampling Task 7A form and submitted to MyPISA.** By defining the initial national desired target population on the form for Sampling Task 7A, and describing any limitations in national coverage, this information will be properly documented and can be formally and adequately presented in the international reports.

Sampling Task 7A form

NPMs should describe their initial national desired target population by specifying enrolment statistics for PISA-eligible students. **Note that although PISA-eligible students include mostly 15-year-olds as well as some students who are 16-year-old, the best information that most countries will have available at the time of completion of Sampling Task 7A is population and enrolment statistics about 15-year-olds. Thus, this is the terminology used on the Sampling Task 7A form.** NPMs are requested to describe the extent of national coverage in their national desired target population. Any populations excluded from the full national population of 15-year-olds need to be documented on this form.

See the template of the Sampling Task 7A form in Appendix D. The Excel template may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/ Sampling Task 7AB.xlsx (the first sheet contains the ST7A form).

Sampling Task 7A form - Example

The following example is based on a fictitious Atlanta plan for PISA for the year 2012 MS. Using information from the *2010 Database of the Statistical Yearbook of Atlanta (school year 2010/2011)*, the NPM found the total national population of 15-year-olds to be 108967. Of this total, 106621 15-year-olds in grades 7 and higher were noted as enrolled in educational institutions.

The NPM reported in question 2.2 that no 15-year-olds were enrolled in grades 6 and below. Of those students in grades 7 and above, no subpopulation of PISA students were to be excluded. Therefore, total enrolment of Atlanta's desired target population was all 106621 15-year-old students. Finally, the NPM noted the source used to define the desired target population and provided a web address where the tables used to complete this form are located.

See the example of the form for Sampling Task 7A in Appendix C.

7.1.2 *Sampling Task Seven B (ST7B): National Defined Target Population*

60. Using their national desired target population as a general framework, NPMs should then describe their national defined target population (or final population). **You will need to submit this information by uploading documents to MyPISA for approval by Westat.**
61. The final national defined target population is essentially reflected by the school sampling frame. Differences between the national desired (initial) and defined (final) target populations will generally arise out of practical considerations and should be kept to a minimum of at most five percent of the national desired target population. **This five percent value though, should not be treated as a limit below which anything is completely acceptable. All exclusions need to be agreed upon.**
62. Usually, practical reasons are invoked for excluding schools and students, such as increased survey costs, increased complexity in the sample design and/or difficult test conditions. Exclusions can occur at the school level, i.e., entire schools are excluded, or within schools, i.e., specific students within sampled schools are excluded. **All such exclusions, at both the school level, and the within-school level, need to be described and quantified on the Sampling Task 7B form and submitted to MyPISA.** This information provides another indication of the quality of the national PISA sample.
63. In PISA, exclusions from the initial national desired target population are to be kept to a minimum. Unlike some other international education studies, difficulties raised by the existence of small schools, (i.e., those with few PISA-eligible students), or other schools where it is difficult or costly to conduct assessments, are mainly to be addressed by modifying the sample design to reduce the number of such schools selected, rather than by exclusion. Similarly, the exclusion of special education students and students with insufficient assessment language experience³ is to be kept to a minimum.

³ Refer to the Technical Standards for PISA 2012 (see MyPISA >Materials > 2012 FT Resources > Cycle Wide > Tech_Standards_PISA12_1.pdf).

7.1.2.1 *School-Level Exclusions*

64. NPMs may wish to exclude a very small fraction of eligible schools for the following reasons:
1. the school is geographically inaccessible (but not part of a region that is excluded from the National Desired Target Population); or
 2. administration of the PISA assessment within the school would not be feasible (e.g., schools for criminals).
65. Exclusion guidelines in the Standards suggest that the percentage of the target PISA-eligible population enrolled in such excluded schools, in total, should be less than 0.5 percent of the population of PISA-eligible students.
66. Note that extremely small schools (i.e., those with exactly only one or two PISA-eligible students) may be excluded **in the field** after initial contact **IF** the number of students in these schools on the frame accounts for no more than 0.5 percent of the final population of PISA-eligible students, **AND** there is still room left in the 0.5 percent exclusion guideline limit, **AND** the Consortium first approves it, **AND** they still only have one or two PISA students at the time of contact, **AND** this will not cause an entire stratum to be missing from the student data. These schools should not be excluded outright from the frame.
67. The NPMs need to state their intentions of excluding these extremely small schools with only one or two PISA-eligible students in the field by answering Question 4 on the Sampling Task 7B form. If agreed upon, these schools should NOT be excluded from the school frame, but if sampled and excluded in the field, their PISA students, weighted by the school weight, will be included in the overall exclusion calculation. Otherwise, the Consortium will request the NPMs to revise their answer to “No” for Question 4. This method is adopted to guard against erroneous size measures leading to more than expected excluded students.
68. In addition, a school may be excluded if it provides instruction only to students in the excluded categories defined under “within-school exclusions” (described in the next section), such as schools for the blind. The guidelines in the Standards indicate that the percentage of PISA-eligible students enrolled in such schools should be less than 2 percent to be eligible for school-level exclusion.
69. Note that if schools of type x are excluded on ST7A, but some such schools are left on the frame, sampled, and found and excluded in the field, the weights of such schools can increase school-level exclusions above acceptable levels. Therefore, be very sure the school frame is clean. If schools of type x are being excluded, make sure ALL type x schools are removed from the school frame, and placed on the exclusion sheet of the school frame.

New for 2012

7.1.2.2 *Within-School Exclusions*

70. NPMs should first note the distinction to be made between within-school exclusions and non-response. Sampled students unable to perform the tests because of a permanent condition (such as those in paragraph 72) are those eligible for exclusion. Students with a temporary condition at the time of testing, such as a broken arm, will be treated as non-respondents along with other “absent” or refusing sampled students.
71. Within schools, all PISA-eligible students should first be listed. Students who would be excluded if sampled need to be retained on the list for sampling, and a variable maintained to briefly describe the reason for exclusion (see DMM Manual). Using this method, the size of the within-school exclusions can be well estimated from the sample data.
72. Because definitions of within-school exclusions will vary from country to country, NPMs are asked to adapt the following rules so that they are workable in their country and coded according to the PISA international coding scheme. These adaptations are negotiated via the country’s Manuals Adaptation Spreadsheet (MAS). International within-school exclusion rules are specified as follows. These refer to SEN codes 1, 2 and 3 respectively as described in the DMM.
1. Students who are functionally disabled in such a way that they cannot take the PISA test. Such students are those with a moderate to severe permanent physical disability. Students with functional disabilities who can respond should be included in the testing.
 2. Students who have a cognitive, behavioural or emotional disability confirmed by qualified staff, meaning they cannot take the PISA test. These are students who are cognitively, behaviourally or emotionally unable to follow even the general instructions of the assessment. Students should not be excluded under this category solely because of poor academic performance or normal discipline problems.
 3. Students with insufficient assessment language experience to take the PISA test. These are students who meet all of the following criteria: a) are not native speakers in the assessment language; b) have limited proficiency in the assessment language; c) have received less than one year of instruction in the assessment language. Only students meeting all three of these criteria can be excluded for this reason.
73. As in previous PISA cycles, there is another within-school exclusion category, referred to as “Other” on the Sampling Task 7B form (and SEN code 4 in the DMM). Although it is expected that the three defined categories should cover all types of within-school exclusions, it may be the case that a particular country needs an additional category. Only a single within-school exclusion category for type "Other" should be used within a country, and this exclusion category should be limited to special circumstances as defined by the NPM and approved by the Consortium before use. Note that if used, the category must be defined

clearly by the NPM so that it is understood in the field. The proposed definition is negotiated via the MAS. (A category defined simply as ‘other’ is not acceptable.)

74. This “Other” category has previously been used for students with dyslexia, dyscalculia or, dysgraphia, as some examples.
75. This “Other” category has previously also been used for country sensitivities where students were not allowed by national law to be classified as within-school exclusion type 1 in one country, or within-school exclusion type 2 in another country. Instead, this “Other” category was used, and the Consortium recoded these SEN code 4 cases to SEN code 1 or SEN code 2 as specified by the country. In such a case, there should be no students having within-school exclusion category 4 on the country’s final student dataset. On the ST7B, these cases should be reflected only through the estimates for within-school exclusion type 1 or type 2.
76. An additional within-school exclusion category is to be used only when a sampled student needs to be excluded because there are no assessment materials available in the student’s Mathematics language of instruction (SEN code 5 in the DMM). Note that this is NOT the same as international within-school exclusion category 3 (paragraph 72) and should never be confused with that category. This code should only be used if it has been suggested by the Consortium.
77. The exact extent of within-school exclusions will not be known until the within-school student sampling frames have been returned from the participating schools and sampling weights computed. **Estimates of the extent of within-school exclusions are required from the NPMs to ensure that the national defined (final) target population will cover at least 95 percent of the national desired (initial) target population. At the time of frame creation, NPMs are asked to provide their best estimate of the extent of within-school exclusions by type of exclusion and report the estimates on the Sampling Task 7B form.** Appendix J provides some possible guidelines for countries which participated in PISA 2009 if no other more up-to-date data is available for these estimates.
78. Countries with adjudicated regions need to submit a ST7A and ST7B, via email to Westat, for each adjudicated region.

Sampling Task 7B form

NPMs need to describe their final national defined target population by describing all types of school-level and within-school exclusions. The estimated size of these excluded populations need also to be reported.

See the template of the Sampling Task 7B Form in Appendix D. The Excel template may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/ Sampling Task 7AB.xlsx (the second sheet contains the ST7B form).

Sampling Task 7B form – Example

The following fictitious discussion uses Atlanta as an example. The national desired target population was composed of 106621 15-year-old students. After excluding 75 SEN schools, the coverage of the national desired target population was reduced by 1.1 percent.

In this example, within-school exclusions of all types were expected to be rare, and therefore deemed as zero. After exclusions, the national defined target population covers 98.9 percent of the national desired target population.

Each type of school-level exclusion is numbered. In this example, school-level exclusion type 1 corresponds to "Special schools for students with special education needs." Section 7.7 illustrates how this exclusion type number is to be used for sample participation status coding.

See the example of the Sampling Task 7B form in Appendix C.

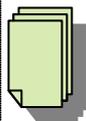
79. Table 1 presents a summary of exclusion limit guidelines.

TABLE 1 EXCLUSION LIMITS SUMMARY

<i>Exclusion Type</i>	<i>Reason</i>	<i>Upper Limit Guidelines</i>
From the International Desired Target Population	Political, organizational, operational	Not specified – should be discussion with and approval by the Consortium. Specify on Sampling Task 7A form.
From the National Desired Target Population:		
School-Level:	Geographical inaccessibility;	0.5% of National Desired

	Administration of PISA not feasible	Target Population. Specify on Sampling Task 7B form.
	All students of types allowed for within-school exclusions if whole school contains only such students	2% of National Desired Target Population. Specify on Sampling Task 7B form.
Within-School Level:	<ul style="list-style-type: none"> -Functionally disabled students; -Cognitively, behaviorally or emotionally disabled students; -Students with insufficient assessment language experience; -Other SEN; -No materials available in student's language of assessment 	<i>Estimate</i> expected to be within 2.5% of National Desired Target Population Specify on Sampling Task 7B form.
TOTAL	-	5% - National Defined Target Population should cover 95% or more of the National Desired Target Population

7.2 Sampling Task Eight: Submitting Your Sampling Frame and School-Level Exclusions



Due two months before the negotiated MS school sample delivery date, or as agreed

80. Like task seven, sampling task eight also involves two subtasks. Subtask A requires you to submit the description of the MS sampling frame using the Sampling Task 8A form (MS Frame Information). Subtask B requires you to submit the full sampling frame and the excluded schools as Sampling Task 8B (MS Sampling Frame and School Exclusion). **You will need to submit the Sampling Task 8A form and the school frame as Sampling Task 8B, as upload documents to MyPISA for approval by Westat.**

New for 2012

81. **To meet all delivery dates, this information must be received by the due date for the negotiated sampling schedule. Late submissions of the set of sampling task forms (ST7A, ST7B, ST8B, and ST8B) means that the negotiated delivery date cannot be guaranteed. With any late deliveries, you should expect the delivery date to be extended by one day for each day the submission is delayed. Similarly, if the sampling frame (ST8B) is of poor quality and it takes x days to improve it, x days will be added to the delivery date.**

7.2.1 Sampling Task Eight A (ST8A): Description of School Sampling Frame

82. Task eight requires all NPMs to construct a school sampling frame covering their national defined target population. School-level sampling frames are based on a comprehensive national list of schools. Area-level sampling frames are used when a comprehensive list of schools is not available or cannot be constructed without undue burden (see Appendix H). In some special cases, NPMs may wish to use a sampling frame that is not related to either of these types. All such cases are required to be approved by Westat. **This type of information about the type of sampling frame that will be used needs to be recorded on the Sampling Task 8A form and submitted to MyPISA.**

83. Other information needed on the Sampling Task 8A form includes what will be used for each school to provide an approximate enrollment (ENR) of PISA-eligible students and the year this data is based on (see also paragraphs 87-90). Similar information is needed for each area on the area-level sampling frame if that is the type of frame used.

84. As a final check of the educational institutions that should appear on the sampling frame, NPMs are asked to indicate in question 8 of the Sampling Task 8A form, for each school type,

whether the frame includes such schools, excludes such schools, or the country has no such schools with the potential to have PISA students at the time of assessment.

Sampling Task 8A form

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

If an area-level sampling frame, or any other type, is to be used, NPMs need to provide a description of the information available to construct this frame. For an area-level sampling frame, this information should be a description of PSUs and their total number, the PSU measure of size available and how many PSUs they intend to sample.

NPMs should also describe the school estimate of PISA enrolment they intend to use.

See the template of the Sampling Task 8A form in Appendix D. The Excel template may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/ Sampling Task 8A.xlsx.

Sampling Task 8A form – School-Level Sampling Frame Description Example

Consider a fictitious example using Atlanta. A comprehensive national list of schools was available and therefore a school-level sampling frame was used. The school estimate of enrolment consisted of enrolment of 15-year-olds (age-based). Enrolment data from the 2010-2011 school year was available for conducting PISA in the year 2012.

Question 8 indicates that all school types that could have PISA students at the time of assessment were included on the frame except for ISCED2 only schools and pre-vocational schools, since there are no such schools in Atlanta, and schools with only SEN enrolment which have been excluded and listed in the excluded schools sheet (sheet 2) of the frame file.

This information is reported in the example of Sampling Task 8A form in Appendix C.

7.2.2 Sampling Task Eight B (ST8B): Sampling Frame and School Exclusion

7.2.2.1 School-Level Sampling Frame

85. A well-constructed school sampling frame is one that provides complete coverage of the national defined target population (final population) without being contaminated by incorrect entries, duplicate entries, incomplete entries, or entries that refer to elements that are not part of the defined target population. Initially, this list should include any school that could have PISA-eligible students at the time of data collection, even those that may later be excluded (see section 7.1.2.1). Additionally, the quality of the data required on the sampling frame has a direct effect on the survey results. Thus frame construction should be treated with extreme care.

86. School-level sampling frames are based on a comprehensive national list of **schools that could contain PISA-eligible students at the time of data collection**. That being said, how is a “school” defined? Although there is no easy answer to this, we generally wish to sample

whole schools as the first stage units of selection, rather than programmes or tracks or shifts within schools, so that the meaning of “between school variance” is more comparable across countries. Sampling units to be used on school-level frames have already been discussed with each country at the time of ST2 completion. See also Appendix I.

87. An approximate enrolment (ENR) of PISA-eligible students (i.e., 15-year-olds) is required for each school listed in the sampling frame. **A suitable school ENR value is a critical component of a school sampling frame since school selection probabilities are based on this quantity.** The quality of a sampling frame will, to a large extent then, depend on the accuracy of the ENR available. The best school ENR for PISA would be the number of PISA-eligible students currently enrolled at the time of data collection. Current enrolment data, however, is rarely available at the time schools are sampled.
88. Suitable measures of the approximate ENR of PISA-eligible students are listed below. In order of suitability, examples of this are:
- Student enrolment of 15-year-olds;
 - If 15-year-olds tend to be enrolled in two or more grades, and the proportions of students who are 15 in each grade are approximately known, the 15-year-old enrolment can be estimated by applying these proportions to the corresponding grade level enrolments;
 - The grade enrolment of the modal grade for 15-year-olds; and
 - Total student enrolment, divided by the number of grades in the school.
89. Note that “student enrolment of 15-year-olds” can be defined in several ways. The first way is the actual number of 15-year-old students (i.e., age-based) in each school on the school frame. The second way, and more care is needed here, is the number of PISA students based on birth dates. There needs to be great care to ensure that if the school data base is based on the school year 2009/2010, for example, and the country is using the standard birth date definition of “born in 1996” for their MS birth date population with spring PISA testing, then it is students born in 1994 that should be used as the ENR estimate. If the school data base is based on the school year 2010/2011, it is students born in 1995 that needs to be used as the ENR estimate. If the school data base is based on the school year 2011/2012, the school year of the 2012 PISA assessment, it is students born in 1996 that should be used as the ENR estimate.
90. If a country knows that the population of 15-year-olds is changing rapidly in the same direction from one year to the next, and if it is previous year frame data that is being used (i.e., 2011 or 2010/2011), it may be better to use for ENR estimates, the number of 14-year-olds in each school. The goal is to have the ENR as close as possible to the actual number of students that will be listed for each school when it is time for student sampling. If this

approach is used, care must be taken if there are ISCED2 only schools or ISCED3 only schools. One strategy may be needed for ISCED2 only schools and another may be needed for ISCED3 only schools.

91. **The exact definition being used for ENR needs to be noted on the Sampling Task 8A form.**
92. Up-to-date and accurate ENR data can greatly improve the precision of the sample drawn. If a country has difficulty in obtaining reasonably accurate estimates of the ENR of 15-year-olds, that is, none of the types of ENR listed in paragraph 88 are available, or the available enrolment data are too out of date, countries are encouraged to discuss with Westat possible ways to improve ENR estimates as soon as this becomes evident. If there is not much improvement that can be made, schools may have to be selected with equal probabilities, which may in turn require an increased school sample size.
93. Note that even the best estimate of ENR may vary from the actual number of PISA students in schools when student lists are collected. The actual number of PISA students in each school based on the school lists will be captured as a separate sampling variable within KeyQuest. KeyQuest will then invoke a comparison between the estimated frame ENR value and the actual PISA students for every sampled and participating PISA school. The schools output from this check are schools where the **difference between these two numbers needs to be investigated. In such cases, schools should be contacted** to ensure that no PISA students were mistakenly left off lists, and/or that extra non-PISA students were not included on the lists by mistake. In this same validity check, KeyQuest will also list out schools where there was no difference between the estimated frame ENR value and the actual number of PISA students in the school since some minor differences are expected even when school frame estimates for ENR are very current. This may indicate students were mistakenly listed only up to the number estimated by ENR, so **lack of any difference should also be queried** to ensure no PISA students were mistakenly left off lists.

New for 2012

94. Each school entry on the frame should include, in addition to a **numeric formatted** estimate of eligible PISA student enrolment:

- School identification information, such as a unique national school ID, and contact information such as name, address, phone number, etc. Note that the Consortium only needs the unique national school ID on the school list that we receive. National Centres need the rest of this information on their full school frame for contacting schools.
- **This unique national school ID must be in text format and needs to be the same length for every school.**

New for 2012

- If you have no unique national school ID and just use sequentially numbered schools as the national school ID values, please add “NATID” as a prefix to this national ID value for each school (e.g., NATID001 for the first school in the frame, NATID002 for the second school in the frame, etc.).

New for 2012

- Information, **all in text format**, about the school that will be used during sampling, such as region of country, school type (public or private), urban/rural classification, school ISCED level (ISCED2, ISCED3, mixed ISCED), etc., which may have some effect on the assessment outcomes, and which were agreed upon in ST2 to be used as stratification variables, or other variables which are required for meeting national option requirements. If the values of these variables are coded (e.g., 01, 02, etc., or 001, 002, etc.), **values should be the same length** for each school.
- If countries are not confident about which strata schools belong to, the school sample size should be increased to offset the negative impact of mis-stratification.

New for 2012

95. Countries may include additional variables on the school frame as they desire. **The format of these variables should be the general format.**

96. The template for ST8B in Appendices C and D show that the structure for the frame is standardized, new for 2012, so that the variables are: National School ID, PISA enrollment, Enrollment of additional population or of PISA subpopulation as defined on ST1, first Explicit stratification variable... sixth Explicit stratification variables, first Implicit stratification variable, ... sixth Implicit stratification variable, extra variable 1 (renamed by country), extra variable n (renamed by country), and **in this specified order.**

New for 2012

New for 2012

97. Countries doing sampling of additional population(s) as a national option (see Appendix G) need to include the estimate of enrolment for each additional population for each school on the school frame. This estimate will be similar to the ENR variable for PISA students, and contains the estimated school enrolment of the additional population. A separate estimate of enrolment for each additional population indicated in the agreed ST1 on MyPISA should be included on the school frame.

98. **The school frame needs to be submitted to MyPISA as Sampling Task 8B and in the form of one Excel file with two data sheets.**

7.2.2.2 Area-Level Sampling Frame

99. As noted previously, area-level sampling frames are used when a comprehensive national list of schools is not available or cannot be constructed without undue burden. As a consequence, area-level sampling frames introduce an additional stage of frame creation and sampling (called the first stage of sampling) before actually sampling schools (the second stage of sampling). This manual provides only an overview of how to implement this additional stage of sampling. See Appendix H. However, the school sampling methods presented in Appendix B can be adapted for that purpose. Furthermore, Westat will provide support to NPMs that

require the use of an area-level sampling frame. **NPMs using an area-level sampling frame should submit to MyPISA, Sampling Task 7AB forms, and a Sampling Task 8A form for the full national defined population. Additionally, they should send via email to Westat, the same for the population numbers corresponding to the sampled first-stage units.**

100. The use of an area-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. Area-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 school-level sampling frame.

7.2.2.3 School-Level Exclusions

101. Based on the National Defined Target Population information reported on the Sampling Task 7B form, some schools may need to be excluded from the school sampling frame. **These schools and their enrolment sizes of 15-year-olds need to be identified on sheet 2 of the Sampling Task 8B school sampling frame, along with the reason for their exclusion. Sheet 1 in this file should be the school sampling frame minus school-level exclusions.**

Sampling Task 8B

NPMs need to submit their sampling frame and exclusions as Sampling Task 8B. It should be submitted as a single electronic Excel file with all the relevant information included. Once the school sample is selected by ACER, the processed sampling frame will be sent back to the NPM with sampled and replacement schools identified and PISA IDs assigned. (See B.8, B.9, and B.10).

See the template of the sampling frame that needs to be submitted as Sampling Task 8B in Appendix D. An Excel template may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/ Sampling Task 8B.xlsx.

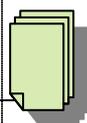
Sampling Task 8B – School-Level Sampling Frame and Exclusions Example

Appendix C contains an example of a sampling frame, again based on the fictitious scenario presented in the previous examples. Only schools from the first few strata are presented for the sake of illustration.

Sheet 2 shows the schools being excluded from the sampling frame.

All details about this example of Sampling Task 8B are found in Appendix C.

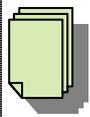
7.3 Sampling Task Nine: Finalize Sample Size Requirements



By one month before the negotiated MS school sample delivery date, or as agreed

102. **No sampling form submissions are necessary for this MS sampling task. The Consortium and the NPMs will discuss and finalize the school sample size and allocation to explicit strata.** See Appendix A for discussion about small schools and their impact on the sample size. See Appendix B for the discussion on how schools are allocated over explicit strata. See Appendix C for completed examples of small school analyses (SSA), and of both the initial and final sample allocation of schools across explicit strata.
103. Countries should plan for adequate time to check the Consortium frame processing, as well as for finalizing the school sample size.

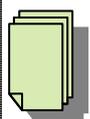
7.4 Sampling Task Ten: MS School Sample Delivery



Due by the negotiated MS school sample delivery date

104. ~~**No sampling form submissions are necessary for this MS sampling task. The NPMs will receive and should review the MS school sample from the Consortium.**~~ The sample selection (see Appendix B) was accomplished by numbering all explicit strata, assigning MOS, producing sampling frame statistics, sorting the explicit strata sampling frames, calculating sampling intervals, creating certainty strata as needed, generating random starts and line selection numbers, determining which schools to sample, and identifying sampled schools and their replacements.
105. See Appendix C for a completed example of the population counts spreadsheet which is produced, and also an example of the selection number Excel file which is produced.

7.5 Sampling Task Eleven: Reviewing and Agreeing to SFKQ



Due one month after the negotiated MS school sample delivery date

106. **The Consortium will create and submit the Sampling Task 11 (ST11) form to the MyPISA website for NPMs to carefully review and agree upon.** This form will contain the following Table 2 variables for each sampled school and will be used as input to create your SFKQ for the MS. These are generally the same variables which were on the FT ST5 which became the SFKQ for the FT. Note that NPMs will have no MS sampling task that corresponds to the FT ST4, which was subsequently uploaded to create FT ST5 (SFKQ). The MS equivalent of FT ST4 will be an Excel sheet that is produced during the MS school

sampling process. Once the final MS school sample (ST10) has been checked by the NPM, ACER will then upload the MS ST11 form for the NPM. The NPM should download the file for review.

107. In Table 2, below, index i refers to additional populations ($i=1$ refers to the first additional population in the Options and Variations form, $i=2$ to the second additional population, etc.), index j refers to sampling steps ($j=1$ refers to the first sampling step, $j=2$ to the second sampling step, etc.), and index k refers to the number of subsamples being selected ($k=1$ means the first subsample being selected, $k=2$ means the second subsample being selected, etc.).

TABLE 2 ALL SFKQ VARIABLES, AND FORMATS

Variable Name	Variable Label	Validation Rule and Value Labels
NC	National Centre 6-digit Code	Like "[0-1][0-9][0-9][0-9][0-9][0-9]"
REGION	Regional Code	Like "[0-8][0-9]"
StIDStrt	PISA Stratum ID	Like "[0-9][0-9]"
StIDSch	PISA School ID	Like "[0-9][0-9][0-9]"
WDUALID	Dual IDs	Like "[3-9][0-9][0-9]" or "000"
WORIGSTRT	Original Stratum	Between 1 and 99 (Equals STIDSTRT if no certainty strata, and the original sampling stratum otherwise)
WNATID	National ID	No Validation Rule; allow up to 50 characters
WSELNUM	Selection Number	Between 1 and 9999 (The sequential number of the sampled school over all strata, from 1 to 9999)
WSCHREP	School Replacement Status	"0" or "1" or "2" (0-Main, 1-First replacement, 2-Second replacement)
WSMOP	Sampling Option	Between 0 and 96
WUH	UH Option	Between "0" and "30" (0-No UH Booklet, 1-UH Booklet in SEN schools, 2-Special agreed schools; agreed with Westat, 3-UH Booklet in General Schools, 4 – FL + UH in SEN schools, 5 – FL + UH in special agreed schools, 6 – FL + UH in general schools)
WEOPTk(k=1,...,4)	Subsample k Optional Component	Between "0" and "9" (0-No Subsamples, 1-Problem Solving, 3-PS+CBAL, 4- PS no UH, 6 – PS+CBAL, no UH, 8- Reading Components)
WGROUP	Instrument Allocation Group	Like "[1-9]"
WENR0	ENR for PISA population	Between 0 and 9900 or 9997 or 9999
WENRi(i=1,...,5)	ENR for Population i	Between 0 and 9900 or 9997 or 9999
WCSj(j=1,...,5)	LST Cap on PSj	Between 0 and 99996, but generally not used
WPSj(j=1,...,5)	Proportional Size j	Between 0 and 1, but generally not used

Variable Name	Variable Label	Validation Rule and Value Labels
WTCSj(j=1,...,5)	Target Cluster Size j	Between -4 and 500 or 99996 or -9
WSGj(j=1,...,5)	Sampling Groups - Set j	Not Like "*[!0-9 OR "", ""]*"
WINDj(j=1,...,5)	Sampling Index j	Like "[0-1][0-1][0-1][0-1]"
WECSk(k=1,...,4)	Subsample k LST Cap on PS	Between 0 and 99996, but generally not used
WEPSk(k=1,...,4)	Subsample k Proportional Size	Between 0 and 1, but generally not used
WETCSk(k=1,...,4)	Subsample k Target Cluster Size	Between -4 and 500 or 99996 or -9
WESGk(k=1,...,4)	Subsample k Sampling Groups	Not Like "*[!0-9 OR "", ""]*"
WEINDk(k=1,...,4)	SUB-Sampling Index k	Like "[0-2][0-2][0-2][0-2]"

108. As noted, variables on MS ST11 are the same as those seen on the FT ST5 (i.e., there are no new variables). However, how some variables are created and used during the MS can differ significantly from how they were created and used in the FT, if they were variables not needed for the FT (e.g., WORIGSTRT, WSELNUM, etc.). All variables will be created during MS school sampling and several will be extracted from your school sampling frame (ST8B) after sampling has been completed.

109. The variables REGION, WNATID, WENR0, WENR_i (i = 1 to 5), WSMOP, WUH, WTCS_j (j=1 to 5), WETCS_k (k=1 to 4), WEOPT_k (k = 1 to 4), and WGROUP, are based on NPM-provided information from several sources including the online Options and Variations form, the online ST1 and ST2 forms for the MS particulars, and the processed ST8B school frame.

110. The variables NC, STIDSTRT, STIDSCH, WDUALID, WSELNUM, WORIGSTRT, and WSCHREP are created during school sample selection. NC is an ACER-defined variable for data management. STIDSTRT, STIDSCH, and WDUALID, are explained in Appendix B, sections B.9 and B.10. WSELNUM and WSCHREP are described in Table 2.

111. The variables WCS_j (j=1 to 5), and WPS_j (j=1 to 5) are generally not used. The variables WSG_j (j=1 to 5), and WIND_j (j=1 to 5) are sampling parameters created by ACER for each student sampling step j.

112. The variables WECS_k (k=1 to 4), and WEPS_k (k=1 to 4) are generally not used. The variables WESG_k (k=1 to 4), and WEIND_k (k=1 to 4) are sampling parameters created by ACER for each student sub-sampling step k.

113. **It is of the utmost importance that you fully understand and agree with all that is in this form.** At the time that you are reviewing this form, if anything is unclear, do not hesitate

to contact Westat for clarification. This form will act as an input file to KeyQuest for student sampling (see Section 7.6.2) and initial validity checks of the submitted data.

7.6 Main Survey Student Sampling Tasks

7.6.1 *Obtaining Student Lists from Participating Schools*

114. After the key components of your manuals have been internationally verified, and about eight weeks prior to the start of testing, and not much before that, you will need to provide instructions to School Coordinators regarding the collection of student lists. Since it is very important that the student sample represent current enrolments, the lists should not be prepared too far in advance.
115. **These instructions appear in the final School Coordinator manual**, and you will need to record adaptations to these instructions as part of the process of recording manual adaptations. The individuals who prepare the lists should also be instructed to return the lists to you at least 6 weeks before testing is to begin. This will provide you with adequate time to select the student samples and send the Student Tracking Form (STF) that lists the selected students back to schools at least 2 weeks prior to testing. This schedule may be affected by school vacations and holidays. You should be aware of the impact of these events and modify the schedule accordingly
116. **Please note the following:** The instructions accompanying the form should convey the following:
- The importance of listing all eligible students, including those who may not be tested due to a disability **or limited language experience**.
 - **The fact that any exclusion from the assessment of students who cannot be tested are required to be done AFTER** the student sample is selected.
 - A definition of PISA-eligible students (in grades 7 or higher born within the population birth date definition that has been agreed with Westat and is recorded on MyPISA).
117. The content of **the form should include the** following (See the DMM for additional information):
- Student line number, grade, gender, month and year of birth, study program and SEN code. If the population definition begins on day 2 or later in the birth month, then a column for day of birth should be included.
 - While it is assumed that the lists of students will contain names, these are not critical to the sampling process as long as the lists contain a unique student identifier. A student identification number assigned by the school, for example, is an acceptable way to identify each student

- There are additional variables that can appear on this form for optional components. If the country is using the UH booklet in general schools, then the list of students needs an extra variable to indicate which SEN students (i.e., those that would otherwise have been excluded) should have the UH booklet. If the country is doing computer based assessments, two other variables for CBA hold the number of available computers, and a comment field about the number of computers.
- If you are selecting an additional grade-based sample and/or other additional population students as a national option (refer to Appendix G) you have to carefully identify on the list of students, those eligible for PISA, and those eligible for grade (or other additional population) sample.
- For this cycle, there is a provision for up to seventeen optional variables on the list of students in KeyQuest. You need to agree with Westat and ACER (via the MAS) on the use of any optional variables. If class sampling is being done, two of these variables need to hold Class name and another for the Class ID that KeyQuest has assigned to each class. If class sampling is not done, the optional variables can be class name, shift, language, or any additional necessary information about students. If used on the list of students, they can be transmitted by KeyQuest into the STF, if desired by the NPM.

118. You will also need to record, as part of the Excel filename for each school's list of students being imported into KeyQuest, the region (REGION), stratum (STIDSTRT), and school ID (STIDSCH) assigned at the time of school sampling and used on your list of participating schools and agreed as part of your sampling plan. **REGION, STIDSTRT and STIDSCH must be used in all parts of your sample tracking and processing**

119. For administrative purposes, you may also need to record additional details such as school name, address, telephone number and email address.

120. We recommend that you ask schools to retain a copy of the list in case you need to call the school with questions.

121. It is important that all students with a date of birth within the PISA population birth date definition enrolled in the school be listed. This includes students who may be excluded from testing because of a special educational need or limited experience in the language(s) in which the test is being offered, students who are frequently absent from school, students in all grades 7 and above offered by the school/sampling unit, and students who may be on work placement programmes at the time of testing. It is very important that the lists of students provided by schools are complete lists of all PISA-eligible students since incomplete lists could seriously compromise the quality of the sample.

122. Once a student list is received from a school, please check it to ensure that all age-eligible students and all relevant information about the students has been received using the following as guidelines.

Check for the number of students: The actual total number of students in the list should be compared to the enrolment variable on the SFKQ file.

If the number is similar, then continue with the next check.

If the number is quite different, for example, smaller by a difference of 10, then the school should be contacted, to query them about the student list. The school should be told that a list with more students was expected and ask them to recheck they have included all eligible students (and explain to them again the target population birth dates). If they have left students off the list, an updated file needs to be requested.

If the school confirms the number of age-eligible students in the list is correct, then continue with the additional checks below.

If the number is quite different, for example, greater by 10, then the school should be contacted to query them about the student list. The school should be told that a list with fewer students was expected and ask them to recheck they have included only eligible students (and explain to them again the target population birth dates). If they have extra non-PISA students on the list that should not be there, an updated file needs to be requested.

If the school confirms the number of age-eligible students in the list is correct, then continue with the additional checks below.

Check student details:

- Does the student list contain details about:
- Year or Grade Level
 - Gender
 - Birth date
 - Etc.

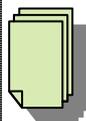
123. If any of this information is missing, or if there are students from only one year level, or only one gender, or students have not been included from the whole target population (for example, missed one month or there are missing levels within gradespan), etc., then send the school a letter, thanking them for sending the student list for PISA, but asking for additional information. Re-attach the list of instructions for preparing the student list, ask them to re-review it, and note that you look forward to soon receiving an updated student list.

7.6.2 Learning the KeyQuest Functions and Operations Required for Student Sampling and Performing Student Sampling in KeyQuest

124. The details of new KeyQuest functions and operations will be provided in the training materials for the NPM meeting in Singapore in October 2011. NPMs are strongly encouraged to send their data managers for this training, especially those considering national or international options related to sampling such as CBA or an additional grade-based population. The final instructions will be available from the DMM.

125. As NPMs and data managers, it will be extremely important to understand these new KeyQuest functions as the quality of student sampling and the weighting of the student data depends on it.

7.7 Sampling Task Twelve: Submitting School Participation Information and Validity Checks to Westat



Due one month after the end of the data collection period

126. You will need to submit your *School Participation instrument (SCP)*, various other KeyQuest instruments (*List of schools, List of students, Student Tracking Form, etc.*), and various KeyQuest validity reports to Westat (STF Internal checks regarding student non-participation codes and SEN code combinations, student non-participation codes and student demographic information, and study programme checks; School participation checks; checks on differences between the school frame ENR values and the schools' actual number of PISA students; checks on CBA and the UH booklet, if either were optional components of the country). See the DMM for further details on the KeyQuest instruments, the validity reports, and data submission.

127. Table 3 summarizes the school participation codes and their meanings. Table 3 also indicates under which circumstances a replacement school should be used.

TABLE 3: PARTICIPATION CODES

Numeric Code for KeyQuest SCP	Code Meaning	Use Replacement?
1	Participant	No
2	Refusal	Yes
3	Non-participation for other reasons	Yes
4	School participated but no students were assessed	Yes
5	No PISA-eligible students	No
6	School closure	No
7	Replacement school not contacted	Not Applicable
8	School participated but nonexcluded sampled students were ineligible	No
9	School exclusion	No
10	School has no PISA-eligible students but has some national option students	No
11	School ineligible for other reasons (please specify)	No

Numeric Code for KeyQuest SCP	Code Meaning	Use Replacement?
12	National Option school to be sampled separately- not needed in 2012	Not Applicable

128. The school participation information needs to be filled out in the SCP instrument in KeyQuest at the end of all sampling processes and submitted to Westat four weeks after the assessment as part of the ST12 data submission task (See the DMM). The following guidelines explain the use of the participation codes in Table 3 under different circumstances.

- If a sampled school’s participation has been confirmed and at least one age-eligible student who was not excluded was successfully assessed, a “1” should be recorded in the School Participation instrument in KeyQuest to denote a participant.
- If a sampled school is unwilling or unable to participate, then enter “2” for refusal, or “3” for non-participation due to other reasons, for instance, school fire, school flood, teacher's strike, etc. A replacement can be used for such a school.
- If the value “3” is used, the reason for non-participation must be entered in the E3COM comment field available in KQ.
- If a sampled school has agreed to participate, but all sampled students who are age-eligible and not excluded refuse to take the test or are absent, NPMs should treat the school as a nonrespondent, and should assign the code ”4” to this school. If time permits, a replacement can be used for such a school.
- If the sampled school is ineligible, then record a “5” for a school with no PISA-eligible students, “6” for a school closure, and “11” for ineligible for other reasons.
- If the value “11” is used, the reason for school ineligibility must be entered in the E3COM comment field available in KQ.
- NPMs may consider using code “11” in the case when a school was mistakenly listed on the frame twice and both records were selected into sample. In this case only one record will be kept. The duplicate record will be made ineligible and assigned a participation code “11” so that the school does not have problems with multiple chances of selection and double coverage. Please contact Westat if this happens to help decide which record to keep and which to call ineligible.
- If a sampled school has agreed to participate, but all sampled students who are not excluded are found to be ineligible (e.g., in grade 6 or lower, birthdates recorded incorrectly and not age-eligible, school transfer after being sampled, drop out, etc), NPMs should enter the code ”8” to this school in the SCP Instrument.
- Furthermore, enter code “9” for a school that belongs to one of the school-level exclusion categories recorded on Sampling Task 7b Question 2, or a school where all age-eligible students are otherwise excluded. Note that a school participation status of “9” requires the school-level exclusion *type* as indicated on ST7b Question 2 to be noted in the E3COM field in the SCP Instrument in KQ, as well as the number of excluded PISA students. Thus, if a sampled school is found to be a school for students with special needs and these schools have been listed as school-level exclusions in Question 2 of ST7b on the line beginning with “1-”, then the school

participation status in SCP will be “9” and “EX-1” should be recorded in the E3COM space along with the number of PISA-eligible students in the school (e.g., EX-1: y students).

- If a school in the sample is found to have all age-eligible students otherwise excluded and no specific school-level exclusion of this type has been previously listed on ST7b (e.g., a school for the blind or more generally a special education school), then use “9” as the status code as usual. For this additional exclusion type, include an exclusion type number in E3COM, “EX-n”, where n is one greater than the number of school level exclusion types noted on ST7b Question 2. Use the same exclusion type number for other sampled schools that are excluded for the same reason. Record also the number of PISA eligible students in the school (e.g., EX-n: y students). Also provide the description for this exclusion type “n”.
- Consider some “9” excluded school examples. Suppose that on the ST7b Question 2 a country excluded language X schools on the line beginning with “1-” and language B schools on the line beginning with “2-”. During data collection, a sampled school is found to have all 10 PISA students instructed in language X. Such a school needs to have a “9” status code and an “EX-1: 10 students” in the E3COM space.
- Further suppose that another school agrees to participate but at the time of assessment, it is determined that all 35 sampled students have been excluded on the basis of the first two within-school exclusion codes. Such a school needs to have a “9” status code and an “EX-3: 35 students” in the E3COM space, with an added description of exclusion type 3 (i.e. students with SEN=1 or SEN=2).
- Suppose once more that a school agrees to participate but at the time of assessment, it is determined that all 30 sampled students have been excluded on the basis of limited language proficiency. Such a school needs to have a “9” status code and an “EX-4: 30 students” in the E3COM space, with an added description of exclusion type 4 (i.e. SEN=3 school).
- If another school agrees to participate but at the time of assessment, it is determined that all 15 sampled students have been excluded on the basis of the fourth “other” within school exclusion codes, such a school needs to have a “9” status code and an “EX-5: 15 students” in the E3COM space, with an added description of exclusion type 5 (i.e. schools with “other” excluded students).
- If another school agrees to participate but all 20 students are instructed in language Y for which there are no test materials, then such a school needs to have a “9” status code and an “EX-6: 20 students” in the E3COM space, with an added description of exclusion type 6 (i.e. no test materials in language of instruction).
- Finally, if a school agrees to participate but all 5 sampled students are excluded for a variety of reasons, then such a school needs to have a “9” status code and an “EX-7: 5 students” in the E3COM space, with an added description of exclusion type 7 (all students excluded for a variety of reasons).
- If a sampled school’s participation has been confirmed and there are no PISA-eligible students in the school (unexpected) but there are national option students, a “10” should be recorded in the School Participation instrument in KeyQuest to denote a participant for the national option only.

- Since all original sampled and replacement schools need to have a final school participation status, all replacements that were not contacted need to have a “7” status code.
- If a country has any schools that have only National Option students in them, by design, then they need to have a final school participation status code of 12. Only one or two countries might need to use this code.
- If a country has students in a school that fit none of these descriptions, contact Westat for help with assigning a school status code.

129. After all contacts with original sampled schools have been made, NPMs will need to contact replacement schools for those sampled schools that will not participate. Each sampled school that will not participate (“02”, “03”, or “04” status codes **only**) should be replaced if possible. **Replacements should NOT be used for reasons other than these. If the original sampled school is ineligible or excluded, do NOT use any replacement schools for it.** Additionally, an original school with any eligible students should not be replaced just because the number of eligible students might be smaller than expected.

130. Second replacement schools should only be used if both the corresponding first replacement school and sampled schools will not participate.

131. As for sampled schools, the participation status of approached replacement schools should be coded as described in paragraph 128.

Appendix A - Treatment of Small Schools

132. For PISA, we identify **small schools, subdivided into moderately small schools, and very small schools**. A **small school** is defined as any school whose approximate PISA enrolment falls below the TCS. A **very small school** is a small school whose PISA enrolment is less than a half the TCS. A **moderately small school** is a small school whose PISA enrolment is in the range of TCS/2 to TCS. Small schools in the sample can result in a reduced sample size of students for the national sample, below the desired target. Additionally, the sample may contain many small schools, which is an administrative burden. In order to minimise these problems, the small schools in the sampling frame may require special treatment.
133. Information on the importance of small schools will be reported in ST9 and in ST10 on the Small School Analysis sheet in the frame file that is returned to the NPM. **As in PISA 2009, no small school strata will be formed. The SSA is done only to enable the final sample size calculation and to determine if very small schools should be undersampled.** The small school analysis is first done on the full school frame overall, and then also for each explicit stratum. The calculated sample size for an explicit stratum is the number of schools to be sampled from all schools in the explicit stratum. Note that because schools are sampled randomly, the numbers of large, moderately small, and very small schools actually selected may vary from the numbers calculated in the explicit stratum small school analysis.
134. The following explains the guidelines that Westat and ACER will generally follow to determine the final school sample size and whether or not undersampling of small schools is needed.
135. To determine whether very small schools should be undersampled and whether sample sizes need to be increased because of small schools, the following test will be applied. If none of the following conditions are true, then no undersampling of very small schools is needed and no increase in the school sample size is needed.
- If the percentage of students in very small schools ($ENR < TCS/2$) is 1 percent or MORE and the percentage of students in moderately small schools ($TCS/2 < ENR < TCS$) is 4 percent or MORE, then very small schools will be undersampled and the school sample size will be increased.
 - Otherwise, if the percentage of students in very small schools ($ENR < TCS/2$) is 1 percent or MORE, then very small schools will be undersampled and the school sample size will be increased.
 - Otherwise, if the percentage of students in very small schools ($ENR < TCS/2$) is LESS than 1 percent, and the percentage of students in moderately small schools ($TCS/2 < ENR < TCS$) is 4 percent or MORE, then there is no undersampling of very small schools but the school sample size will be increased.

- Otherwise, there will be no adjustments to the school sample size.

136. The following discussion gives the approach that will be used by Westat and ACER to determine the minimum sample size of small schools (and of larger schools), that ensures that the resulting student sample is representative, yet gives some control to limit the total sample size of schools in extreme cases. The ensuing discussion, based on an overall small school analysis on the full school frame, assumes that the school sample size needs to be inflated because of very small and moderately small schools, and that very small schools need to be undersampled. The formulae below also require modification when TCS is other than 35.

New for 2012

137. If undersampling of very small schools is required, then the undersampling will be done in two parts. Schools with enrolment of PISA students equal to zero, one, or two, will be undersampled by a factor of four, and schools with enrolment of PISA students greater than two and less than TCS/2 will be undersampled by a factor of two.

138. The small school analysis is conducted as follows:

- From the complete sampling frame, find the proportions of total ENR that come from very small schools with zero, one or two PISA students (P1), very small schools with more than two PISA students but fewer than TCS/2 PISA students (P2), moderately small schools (Q), and larger schools (those with ENR of at least TCS) (R). Thus $P1+P2+Q+R=1$.

- Calculate the figure L, where

$$L = 1 + 3(P1)/4 + (P2)/2.$$

Thus L is a positive number, a little more than 1.0.

- The minimum sample size for larger schools is equal to $150 \times R \times L$, rounded up to the nearest integer.
- Calculate the mean value of ENR for moderately small schools (MENR), and also for very small schools (V2ENR and V1ENR). Thus MENR is a number in the range of TCS/2 to TCS, V2ENR is a number no greater than TCS/2, and V1ENR is in the range of zero to two.
- The number of moderately small schools that need to be sampled is given by the formula below and then rounded up:

$$(5250 \times Q \times L) / (\text{MENR}).$$

- The number of very small schools of type P2 that need to be sampled is given by the formula below and then rounded up:

$$(2625 \times P2 \times L) / (\text{V2ENR}).$$

- The number of very small schools of type P1 that need to be sampled is given by the formula below and then rounded up:

$$(1313 \times P1 \times L) / (V1ENR).$$

139. Consider the following example. Suppose that in Country X, the TCS is equal to 35, with 0.1 of the total enrolment of 15-year-olds in moderately small schools, with 0.05 in very small schools of type P2, and 0.05 in very small schools of type P1. Because enrolment in very small schools is 1% or more, undersampling of very small schools is required. Suppose that the average enrolment in moderately small schools is 25 students, while the average enrolment in very small schools of type P2 is 12 students, and in very small schools of type P1 is 1.5 students.

- Thus $P2 = 0.05$, $P1 = 0.05$, $Q = 0.1$, $R = 0.8$, $MENR = 25$, $V2ENR = 12$, and $V1ENR = 1.5$.
- From b, $L = 1.0625$.
- Thus, from c, the sample size of larger schools should be at least $150 \times 0.8 \times 1.0625 = 128$. That is, there should be at least 128 of the larger schools sampled.
- From e, the number of moderately small schools required is $(5250 \times 0.1 \times 1.0625) / 25 = 22.3$; i.e., 23 schools.
- From f, the number of very small schools required of type P2 is $(2625 \times 0.05 \times 1.0625) / 12 = 11.6$; i.e., 12 schools.
- From g, the number of very small schools required of type P1 is $(1313 \times 0.05 \times 1.0625) / 1.5 = 46.5$; i.e., 47 schools.
- This gives a total sample size of $128 + 23 + 12 + 47 = 210$ schools. The larger schools will yield an initial (i.e., before considering school and student nonresponse) sample of $128 \times 35 = 4480$ students. The moderately small schools will give an initial sample of approximately $23 \times 25 = 575$ students. The very small schools of type P1 will give an initial sample size of approximately $12 \times 12 = 144$ students. The very small schools of type P2 will give an initial sample size of approximately $47 \times 1.5 = 71$ students. Thus the total initial sample size of students is $4480 + 575 + 144 + 71 = 5270$. Our sample, modified to handle the small schools, will yield approximately the target student sample size that we seek. Although more schools are being sampled with this modification, many of the 47 P1 sampled schools will have no PISA students and will therefore be ineligible so that little effort needs to be spent on those schools.

140. Why this adapted approach? Several countries have a large number of schools with zero, one, or two PISA students, and in PISA 2009 this initially resulted in a very large school sample size. These countries requested undersampling all very small school by a factor of four to alleviate the large sample size. Undersampling all very small schools by a factor of four however, will usually have adverse effects on sampling variability and should be avoided. Thus, this modification is a compromise so that only the very smallest schools may be undersampled by a factor of four and the other very small schools by the more acceptable factor of two.

141. If a country does not have the need to undersample very small schools, then Westat and ACER will set $MOS = TCS$ for all small schools. If a country needs undersampling of very small schools, then this is accomplished during sample selection by Westat and ACER setting $MOS = TCS/2$ for very small schools of type P2 (e.g., schools with size smaller than $TCS/2$ and greater than two), setting $MOS = TCS/4$ for very small schools of type P1 (e.g., schools with sizes zero, one or two PISA students) and setting $MOS = TCS$ for moderately small schools (e.g., schools with size between $TCS/2$ and TCS).

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis

Westat and ACER will use student enrolment data provided by the NPM on the sampling frame to determine the prevalence of small schools in their school system. Based on these enrolment data, specific sample design decisions may be required if the presence of small schools is substantial. If sampling will not be conducted by Westat and ACER, the analysis of small schools is still needed both for the overall country and also by explicit stratum. A worksheet is provided in Appendix D to assist NPMs with the evaluation procedure.

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis – Example

Using a hypothetical scenario for Atlanta's PISA 2012 MS in Appendix C, there are more than 1 percent eligible students in very small schools (enrolment of 15-year-olds less than 18) and more than 4 percent eligible students in moderately small schools (enrolment of 15-year-olds between 18 and 35). Therefore, the overall analysis indicates Atlanta will need to increase the sample size and undersample very small schools. There are no schools in Atlanta's school frame with 0, 1, or 2 students.

The example in Appendix C shows the worksheet for the overall Small School Analysis, the worksheet for the first explicit stratum, and the same information in the Excel layout of the Small School Analysis sent to NPMs for ST9.

An additional table shows the Small School Analysis columns with calculations for the needed school sample sizes. The base number of schools/students for individual strata Small School Analysis follow from the initial sample allocation (see Appendix B). Note that the SSA is performed separately for each explicit stratum by the procedure outlined in paragraphs 134 and 138 with minor adjustment for the base number of schools and students.

Appendix B - School Sampling Selection Methodology

142. The information in this section is provided for information purposes only.

B.1 Strata Numbering

143. If there are explicit strata, they will be numbered sequentially and they will be labeled. Schools selected with certainty will be placed in the certainty stratum after all other strata (See B.7).

B.2 School Sample Allocation Over Explicit Strata

144. If there are explicit strata, the total number of schools to be sampled in each country will be allocated among the explicit strata so that the percentage of students in the sample in each stratum is approximately the same as the percentage of students in the population in each stratum. This sample allocation is done by the Consortium and reported on ST9: PISA 2012 Main Survey School Sample Selection: Sample Allocation by Explicit Strata (a sheet in the frame file that is returned to the NPM).

145. The following points are taken into account.

- If separate estimates are required for explicit strata, then equal levels of sampling precision will usually be required for each explicit stratum. This generally means that there will be an equal number of schools sampled in each explicit stratum regardless of the size of each stratum.
- If different TCSs are applied to the explicit strata, the sample allocation requires adaptation.
- All strata have to have at least one allocated school. Additionally, any explicit stratum that has only one school allocated will instead have at least two allocated to it in the final sample allocation for purposes of variance estimation.
- If an explicit stratum has only one school in the stratum sampling frame which will not be sampled, then all schools in such a stratum will be included in the sample.

ST9: PISA 2012 Main Survey School Sample Selection: Sample Allocation by Explicit Strata

Westat and ACER will describe the country's school sample allocation in ST9: PISA 2012 MS School Sample Selection: Sample Allocation by Explicit Strata, two sheets in the file returned to the NPM, one showing the initial allocation and one the final allocation after Small School Analysis and with sample size adjustments as needed. The final sample allocation will show the allocation of schools to the explicit strata.

ST9: PISA 2012 Main Survey School Sample Selection: Sample Allocation by Explicit Strata - Example

The example depicted in Appendix C involves fictitious information for Atlanta's sample design.

The example of ST9: PISA 2012 MS School Sample Selection: Sample Allocation by Explicit Strata shows what Atlanta's proposed stratification and allocation would be like. Explicit strata are defined based on state, and allocation is done proportionally so that the percentage of eligible students in the sample in each stratum is about the same as the percentage of eligible students in the population in each stratum. Note that the total sample size is initially 150 schools since Atlanta does not anticipate any field school exclusions, ineligibility, or nonresponse. The second sheet shows the needed adjustment to the initial allocation due to the population enrolled in small schools and following from the Small School Analysis.

B.3 Assigning a Measure of Size (MOS) to each School

146. As indicated earlier, for each school on the frame an approximate ENR value for the school is needed. In some cases this will be an actual enrolment of 15-year-olds, based on administrative records for the current year. In many countries the number will be the enrolment in the grade containing the most 15-year-olds, either for the current year or the previous year.

147. For the PPS sampling method to be used for PISA, for each school we need to establish a measure of size (MOS), derived from ENR. This will be done by Westat and ACER, so what follows is for information purposes only. The MOS, assigned as part of the MS school sampling procedure on the Sampling Task 8b frame, is given as: $MOS = \text{MAX}(ENR, TCS)$.

148. That is, the measure of size is equal to the ENR estimate, unless the enrolment is less than the TCS, in which case the measure of size is set equal to the TCS (or $TCS/2$ or $TCS/4$ — See Appendix A).

149. This, and the increased sample size for small schools, if necessary, ensures that the sample has a good representation of students from small and very small schools, as well as large schools (See Appendix A).

B.4 Sampling Frame Statistics

150. The Consortium will tabulate basic statistics by strata from the school sampling frame and send this back to the NPM as ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata (a sheet in the frame file that is returned to the NPM).

151. For each combination of explicit and implicit strata, the number of schools and enrolled students (both ENR and MOS) will be reported. These statistics will be very useful to finalise the sample allocation of schools to the explicit strata and as control totals to validate the computation of sampling weights. Very large schools selected with certainty and removed to a separate stratum at the time of sampling are included in these tabulations in their original strata.

ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata

Westat and ACER will list and describe all combinations of explicit and implicit strata in ST10: PISA 2012 MS School Sample Selection: Population Counts by Strata, on a sheet in the returned sampling frame with the sample selected. For each entry, the number of schools, eligible students, and school MOS will be tabulated. These statistics should be readily available from the school sampling frame.

ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata – Example

Appendix C contains a completed example of ST10: PISA 2012 MS School Sample Selection: Population Counts by Strata based on the fictitious scenario for Atlanta. The example form shows the states of Atlanta as explicit strata and urbanisation as an implicit stratification variable.

A brief note about stratifier level coding: in preparation for sampling, ACER will create a copy of each ST8B frame stratifier and assign a numeric code for each of a stratifier's levels, this is to assist with frame sorting during sampling. To illustrate, State is the first Explicit stratification variable, with values State1, State2, etc. These are recoded as 1,2, etc. Similarly, the two levels of Atlanta's implicit urbanisation stratifier variable ("Urban" and "Rural") seen in a previous example were coded as 1=Urban and 2=Rural. Subsequently, the value "1,1" in the second column of Atlanta's ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata in Appendix C reflects Atlanta's State 1 schools that are in Urban location, and the value "1,2" in this column denotes the State 1 schools that are in Rural location. Tables showing the number code for each ST8B stratifier level will be provided to the NPM with Sampling Task 9 for review, and then also returned to the NPM in the ST10 frame file as part of the sampling process documentation.

B.5 Sorting the Sampling Frame

152. If explicit strata are used, what follows in this section, regarding sorting of the frame, should be done within each explicit stratum. Failing to sort the sampling frame prior to sample selection will not bias or invalidate the original sample. This could increase nonresponse bias though, since replacement schools will not be as similar as possible to their originally sampled nonparticipating schools. Additionally, not sorting by the implicit stratifiers may unnecessarily increase the sampling variances of estimates obtained from the assessments, as the implicit stratification is implemented via this sort process.

153. On ST2, 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.

154. 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 within the levels of the first sorting variable, and so on, until all implicit stratification variables have been exhausted. The result is a cross-classification structure of cells, where each cell represents one implicit stratum on the 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. It is the maximum number because some cross-classifications could contain no schools at all.

155. Within each implicit stratum, the schools are sorted by their ENR values - NOT their MOS values. The sort order should alternate from high to low and then low to high, as indicated in Table 4. For more than four implicit strata, the sort order pattern as presented in Table 4 should simply be continued.

TABLE 4: IMPLICIT STRATUM SORT ORDER

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

B.6 Determining Which Schools to Sample

156. ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers sheet with the selected sample is used to identify which schools are selected in the sampling frame. There should be one of these for each explicit stratum. This is used to generate a series of selection numbers, which will be applied to the sampling frame to identify the sampled schools. The ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers worksheet is provided in Appendix D for reference, but ACER will complete this for each explicit stratum as part of the sampling procedures and will return it to the NPM as a separate file to accompany the frame file in ST10.

B.6.1 Calculating the Sampling Interval

157. The PPS systematic sampling method requires the computation of a sampling interval for each explicit stratum. This calculation is performed based on information recorded in ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers.

158. To calculate the sampling interval:

- The total measure of size “S” for all schools in the sampling frame for the specified explicit stratum is recorded in ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers.
- The number of schools “D” to be sampled from the specified explicit stratum is recorded in ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers. This will be the number allocated to the explicit stratum in the final sample allocation following from ST9 discussion with Westat. The numbers from 1 through “D” are entered sequentially in column 1 (line numbers) of ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers, for later use.
- The sampling interval “I” is calculated as: $I = S \div D$. The sampling interval “I” is recorded to four decimal places in ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers.

B.6.2 Generating a Random Start

159. A random number is generated for each explicit stratum. Random numbers are now easily generated by computer software, such as Excel, SAS or SPSS. These random numbers should be chosen from a “uniform random number generator” that generates numbers distributed uniformly between 0 and 1.

160. The generated random number “RN” is recorded in ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers. The selected random number should be a number between 0 and 1 and should be recorded to four decimal places.

B.6.3 Calculating the Selection Numbers

161. Selection numbers are calculated and recorded on ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers 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”. This first selection number is recorded to four decimal places in column 2 of the first line, corresponding to line number 1, on ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers. This first selection number will be used to identify the first sampled school in the specified explicit stratum.
- We obtain the second selection number by simply adding the sampling interval “I” to the first selection number. The second selection number will be used to identify the second sampled school. This number is recorded to four decimal places on the second line of column 2.
- 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.

162. To summarize the calculation of selection numbers, the first selection number in an explicit stratum is $RN \times I$. The second selection number is $(RN \times I) + I$. The third selection number is $(RN \times I) + I + I$, and so on.

163. Selection numbers need to be generated independently for each explicit stratum, starting with a new random number for each explicit stratum. For each explicit stratum, the first line number in column 1 of ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers is always one.

ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers

ACER will complete ST10: PISA 2012 MS School Sample Selection: Selection Numbers to demonstrate how the school sample was selected. The information and calculations will be returned to the NPM as a

file of Selection Numbers that contains a separate sheet for each explicit stratum.

ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers – Example

The example of ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers (see Appendix C) is based on the third of Atlanta's explicit strata from a previous example.

This scenario requires the selection of 13 schools. This example consists of a single explicit stratum. In “S” we have recorded the total MOS for that stratum (measure of size for state 3 from column 5 of the earlier Atlanta example of ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata). In “D” we have recorded the allocated school sample required: 13. The resulting sampling interval is recorded in “I” and is in fact 744.3846. A random number is generated and recorded in “RN”. The example of ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers lists 13 line numbers with all the generated selection numbers needed to select the sample of 13 schools in the sampling frame for this explicit stratum.

B.7 Certainty School Stratum

164. Some countries will have schools that will be selected with certainty. This means that within their explicit sampling stratum, their MOS values were equal to or greater than the stratum sampling interval (i.e., the total cumulative stratum MOS divided by the number of schools to be sampled from the stratum (see also B.6.1)). In such cases, the certainty selections are removed from their stratum, and placed in a certainty stratum. The sampling interval in that stratum from which the certainty was removed is re-calculated for the selection of the rest of its sample. Note that there may be additional certainty selections for the re-calculated sampling interval and the process of removing these large schools and re-calculating a new interval iterates until there are no further certainty selections. Any iterative calculations of the sampling interval due to certainty selections are provided in a sheet in the frame file returned to the NPM. The stratum to which each certainty selection originally belonged will be apparent from the original stratum identifier WORIGSTRT that will be available on the frame data file and on ST11.

B.8 Identifying the Sampled Schools

165. Based on the selection numbers recorded in ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers, we are now able to identify the sampled schools in 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 corresponding explicit strata in each sampling frame.

166. The first task, however, is to compile a cumulative measure of size in the school sampling frame. This cumulative MOS will determine which schools are sampled. The cumulative MOS is tabulated within explicit strata. This means the cumulative MOS needs to be reset when moving from one explicit stratum to the next.

167. Sampled schools are identified as follows on the sorted Sampling Task 8b frame:

168. Let “Z” denote the first selection number for a particular explicit stratum. We need to 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_s is the cumulative MOS of a particular school in the sampling frame and $C_{(s-1)}$ is the cumulative MOS of the school immediately preceding it, then the school with the cumulative MOS C_s is selected if:

- C_s is greater than or equal to Z, and
- $C_{(s-1)}$ is strictly less than Z.

169. Apply the above rule to all selection numbers for a given explicit stratum and enter “S” for each sampled school on the sorted frame.

B.9 Identifying Replacement Schools

170. NPMs should make every effort to get as many of the sampled schools to participate as possible. However, 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 a priori replacement schools for non-participating sampled schools is needed. A second, perhaps more important, reason for identifying replacement schools a priori is to avoid the haphazard use of alternate schools as replacements, which may actually amplify response biases. Although our approach is no guarantee of avoiding non-response biases, it will at least tend to minimize the potential for bias. The technique of identifying replacements should lead to less non-response bias than the alternative of drawing a larger sample initially, in anticipation of nonresponse. Each sampled school in the main survey will be assigned two replacement schools in the sampling frame if possible. **A sampled school can never be designated as the replacement school for another sampled school.**

171. Replacement schools are identified as follows. For each sampled school, identified by an S in the sorted sampling frame, the schools immediately preceding and following it are designated as its replacement schools. The school immediately following the sampled school should be labelled “R1” (i.e., first replacement), while the school immediately preceding the sampled school should be labelled “R2” (i.e., second replacement). The use of implicit stratification variables, and the subsequent ordering of the school sampling frame by size (ENR), ensures as much as possible that any sampled school’s replacements will have similar characteristics. If it sometimes occurs that the first replacement school does not have the same set of implicit stratification values as the sampled school, the first and second replacement schools may be switched. It may be difficult to assign replacement schools to some very large sampled schools because the sampled schools appear very close to each other in the sampling frame. Therefore, at times, only a single replacement school will be assigned, and perhaps none when two consecutive schools in the sampling frame are sampled.

172. In small countries, there may be problems encountered when trying to identify two replacement schools for each sampled school. In such cases, a replacement school could be allowed to be the potential replacement for two sampled schools (e.g., a first replacement for the first sampled preceding school, and a second replacement for the following sampled school), but an actual replacement for only one school. Such a single replacement school between two sampled schools will be noted with two unique STIDSCH values and with an RD designation as opposed to an R1 or R2. This is called a dual role replacement school. (see paragraph 176)

173. 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. The school immediately above the sampled school is labelled R1, and the school above R1 is labelled R2. Should a sampled school happen to be the first school listed in an explicit stratum, then the two schools immediately below it are its designated replacement schools. The school immediately below the sampled school is labelled R1, and the school below R1 is labelled R2.

B.10 Assigning STIDSTR and STIDSCH

174. In order to keep track of sampled schools and replacement schools in the PISA database, all sampled schools and replacement schools will be assigned unique identification numbers. **The STIDSTR and STIDSCH will be assigned to all sampled schools and replacement schools and recorded on the frame file returned to the NPM. All NC submissions of school or student data after school sampling MUST include the Consortium-assigned STIDSTR and STIDSCH.**

175. School ID's will be assigned according to these rules:

- Schools will be assigned a two-digit stratum code STIDSTR (corresponding to the explicit strata), and a three-digit school code STIDSCH.
- Sampled schools are sequentially numbered starting from 1(one), within each explicit stratum. For example, if 150 schools are sampled from a single explicit stratum, then they are assigned STIDSCH from 001 through 150. This numbering should continue across all implicit strata.
- First replacement schools in the main survey are assigned the STIDSCH of their corresponding sampled schools, but incremented by 300. For example, the first replacement school for sampled school 023 is assigned STIDSCH= 323.
- Second replacement schools in the main survey are assigned the STIDSCH of their corresponding sampled schools, but incremented by 600. For example, the second replacement school for sampled school 136 is assigned STIDSCH= 736.

176. Consider an example for RD schools. Suppose the stratum frame has first and second STIDSCH school identifiers as follows. Here, the RD school is the one with both STIDSCH = 303 and STIDSCH = 604. If school 003 refuses to participate, then the RD school should be approached as a replacement for 003. If however, school 003 participates but school 004 cannot participate because of a teacher strike, then the school with ID of 304 should be

approached first. If school 304 refuses to participate, then the RD school can be approached as a second replacement for school 004. It should be apparent that the RD school can only be used once—either as a first replacement for school 003 or as a second replacement for school 004.

603	
003	
303	604
004	
304	

Appendix C – Examples of Completed Sampling Forms

ST7A: PISA 2012 Main Survey National DESIRED Target Population (EXAMPLE)

See Section 7.1.1 of School Sampling Preparation Manual.

Please complete this form in full before filling out ST7B. Enter information in the cells that are highlighted in green. Make sure number values have no commas or decimal points. Please complete all highlighted sections of this form. After completing this form please do ST7B.

	PISA Participant:	Atlanta
	Person who completed this form:	Mr. X
	Date this version of this form was completed:	15-10-2011
1	Total national population of 15-year-olds:	108967
2.1	Total national population of 15-year-olds enrolled in educational institutions and who are in grades 7 and higher:	106621
2.2	Total number of 15-year-olds enrolled in grades 6 and below:	0
3.1	If the national desired target population for your country differs from the total national population of enrolled 15-year-olds in grades 7 and higher, describe the population(s) to be excluded from the total national population of enrolled 15-year-olds in grades 7 and higher. These include distinct geographic regions of the country, or specific language groups. If no population is to be excluded please enter "None".	None
3.2	Total enrolment <u>excluded</u> from the total national population of enrolled 15-year-olds in grades 7 and higher (corresponding to the exclusions listed above):	0
4	Total enrolment in the national desired target population:	106621
5	Percentage of coverage of nationally enrolled 15-year-olds, in the national desired target population:	100.00%
6	Describe your data source and year of data. Also provide webpage address with relevant tables. Tables not in English should be emailed to Westat with English labels added for relevant table variables:	Database of Statistical Yearbook (school year 2010/2011), www.ATL.education.gov , Statistical Yearbook of Atlanta, 2010
	Please include any additional comments you may have in the box below:	

Thank you for completing ST7A, please continue to ST7B.

End of ST7A

**ST7B: PISA 2012 Main Survey National DEFINED
Target Population (EXAMPLE)**

See Section 7.1.2 of School Sampling Preparation Manual.

Please complete ST7A before you begin to fill out this form. Enter the relevant information in the cells that are highlighted in green. Make sure number values have no commas or decimal points. Please complete all highlighted sections of this form.

PISA Participant: **Atlanta**
 Person who completed this form: **Mr. X**
 Date this version of this form was completed: **15-10-2011**

1 Total enrolment in the national desired target population: **106621**

2.1 School-level exclusions:

Please write a brief description of each type of school-level exclusion in the table below and for each type of exclusion you enter, please include the number of schools and students to whom this exclusion will apply.

	<u>Description of school-level exclusions</u>	<u>Number of schools</u>	<u>Number of students</u>
1	Special schools for students with special needs	75	1162
2			
3			
4			
5			
6			
7			
8			
9			
10			
	TOTAL	75	1162

2.2 Percentage of students not covered due to school-level exclusions: **1.09%**

3 Total enrolment in national defined target population before within-school exclusions: **105459**

4 Do you plan to exclude schools which have only one or two PISA students in the field during data collection and only where this will not cause an entire stratum to be missing from the collected data?

No

5.1 Anticipated within-school exclusions (students who could not be included in the PISA assessment, from schools where some students could be included). Please enter 0 (zero) when no students of the type will be excluded.

<u>Descriptions of within-school exclusions</u>	<u>Expected number of students</u>
Functionally disabled students	0
Intellectually disabled students	0
Students with insufficient language experience in test language	0
Other	0
No materials available in student's language of mathematics instruction	0
TOTAL	0

5.2 Expected percentage of students not covered due to within-school exclusions: **0.00%**

6 Total enrolment in national defined target population: **105459**

7 Coverage of national desired target population: **98.91%**

8 Describe your data source and year of data. Also provide webpage address with relevant tables. Tables not in English should be emailed to Westat with English labels added for relevant table variables:

2010/2011 Statistical report of the Department of Education, Atlanta: Stationary Office, www.ATL.Education.gov

Please include any additional comments you may have in the box below:

Thank you for completing ST7B. You may now submit this form.

End of ST7B

ST8A: PISA 2012 Main Survey Sampling Frame Information (EXAMPLE)

See Section 7.2.1 of School Sampling Preparation Manual.

Enter information in the cells that are highlighted in green. Please complete all highlighted sections of this form, unless otherwise specified. In many of the cells you will need to choose a response from a drop-down list. To do this, first select the cell, then click on the arrow symbol that appears and choose the appropriate response from the list. An error message will appear if an invalid response is selected.

PISA Participant: Atlanta
Person who completed this form: Mr. X
Date this version of this form was completed: 15-10-2011

1 Will a sampling frame of geographic areas be used? No

Please skip ahead to question 5.

Please do not provide an answer to this question.
2 Specify the PSU Measure of Size to be used:

Please do not provide an answer to the following question.
3 Specify the school year for the data which will be used for the PSU Measure of Size:

Please do not provide an answer to the following question.
4 Please provide a preliminary description of the information available to construct the area frame. Please consult with Westat for support and advice in the construction and use of an area-level sampling frame.

Please answer all questions below.
5 Specify the school estimate of enrolment of 15-year-olds that will be used. 15-year-old student enrolment (age-based)

6 Specify the year for which enrolment data will be used for the school estimate of enrolment of 15-year-olds: 2010/2011

- 7 Please describe any other type of frame, if any, that will be used. If no other type of frame will be used please enter "None".

None

- 8 Please indicate whether the following types of schools exist in your system and whether they have been included or excluded in your PISA sampling frame. Each school type requires a response.

<u>School Type</u>	<u>Included / Excluded / Don't Have Any</u>
Bilingual schools	Included
Immersion schools	Included
National schools with single (minor) language of instruction	Included
International / foreign schools with single (minor) language	Included
Schools with mixed (major and minor) languages of instruction	Included
Schools with fewer than 3 estimated PISA students	Included
Non-publically funded schools	Included
Rural schools	Included
Urban schools	Included
Remote schools	Included
ISCED2 only schools	Don't Have Any
ISCED3 only schools	Included
Mixed ISCED schools	Included
Pre-vocational schools	Don't Have Any
Vocational schools	Included
Technical schools	Included
Agriculture schools	Included
Other vocational schools	Included
General schools	Included
Mixed programme schools	Included
Predominantly male schools	Included
Predominantly female schools	Included
All male schools	Included
All female schools	Included
Schools with only part-time students	Included
Schools with multiple shifts	Included
Religious schools	Included
Schools with predominantly international / foreign students	Included
Schools with mixed national / foreign students	Included
SEN only schools	Excluded
Any other irregular schools	Included

Please include any additional comments you may have in the box below:

Thank you for completing ST8A. You may now submit this form.

End of ST8A

**ST8B: PISA 2012 Main Survey Sampling Frame and School Exclusion File:
Sheet 1 (EXAMPLE)**

See Section 7.2.2. The Excel template for Sampling Task 8B may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/

National School ID	PISA enrolment	Enrolment of additional population or of PISA sub-population as defined on ST1	First Explicit Stratification Variable Specified on ST2	Second Explicit Stratification Variable Specified on ST2	...	First Implicit Stratification Variable Specified on ST2	Second Implicit Stratification Variable Specified on ST2	...	Extra Variable 1 (renamed by country)	Extra Variable 2 (renamed by country)	...	Extra Variable n (renamed by country)
2010096	342	0	State 1			Urban						
1000042	281	0	State 1			Urban						
8010040	271	0	State 1			Urban						
1020020	241	0	State 1			Urban						
8010140	234	0	State 1			Urban						
10012532	233	0	State 1			Urban						
1010098	227	0	State 1			Urban						
7010044	224	0	State 1			Urban						
7012686	221	0	State 1			Urban						
4012583	212	0	State 1			Urban						

5010151	198	0	State 1			Urban						
1010063	196	0	State 1			Urban						
8010001	185	0	State 1			Urban						
1010041	184	0	State 1			Urban						
8010022	166	0	State 1			Urban						
2010201	157	0	State 1			Urban						
9010159	156	0	State 1			Urban						
5010141	147	0	State 1			Urban						
3010002	146	0	State 1			Urban						
1010195	144	0	State 1			Urban						
10010069	143	0	State 1			Urban						
5010160	142	0	State 1			Urban						
3012571	134	0	State 1			Urban						
3010087	133	0	State 1			Urban						
5010162	133	0	State 1			Urban						
2010184	133	0	State 1			Urban						
9010188	128	0	State 1			Urban						
7010236	127	0	State 1			Urban						
7010347	126	0	State 1			Urban						
1010202	122	0	State 1			Urban						
0010084	118	0	State 1			Urban						
9010039	116	0	State 1			Urban						
8010015	108	0	State 1			Urban						
2010007	104	0	State 1			Urban						
7010237	104	0	State 1			Urban						

10010193	102	0	State 1			Urban						
6012529	101	0	State 1			Urban						
10010225	100	0	State 1			Urban						
4010060	100	0	State 1			Urban						
1010215	99	0	State 1			Urban						
6010023	96	0	State 1			Urban						
8010213	95	0	State 1			Urban						
6010187	93	0	State 1			Urban						
0010210	90	0	State 1			Urban						
8010002	90	0	State 1			Urban						
7010003	88	0	State 1			Urban						
2010243	86	0	State 1			Urban						
6010185	84	0	State 1			Urban						
3012715	83	0	State 1			Urban						
9010091	83	0	State 1			Urban						
3010070	82	0	State 1			Urban						
5010152	81	0	State 1			Urban						
1010066	76	0	State 1			Urban						
4010045	73	0	State 1			Urban						
6010013	73	0	State 1			Urban						
7010155	69	0	State 1			Urban						
10010085	65	0	State 1			Urban						
0012597	52	0	State 1			Urban						
9010211	46	0	State 1			Urban						
8010082	42	0	State 1			Urban						

5010011	32	0	State 1			Urban						
0010199	24	0	State 1			Urban						
5012614	157	0	State 1			Rural						
1010064	163	0	State 1			Rural						
3010079	202	0	State 1			Rural						
8012625	253	0	State 1			Rural						
4022648	383	0	State 2			Rural						
8020522	316	0	State 2			Rural						
1020541	310	0	State 2			Rural						
3022602	278	0	State 2			Rural						
6022649	243	0	State 2			Rural						
10020112	209	0	State 2			Rural						
1020806	204	0	State 2			Rural						
9021114	203	0	State 2			Rural						
0020728	182	0	State 2			Rural						
7022579	181	0	State 2			Rural						
10022654	176	0	State 2			Rural						
1020701	168	0	State 2			Rural						
9022719	165	0	State 2			Rural						
3022658	164	0	State 2			Rural						
4020802	161	0	State 2			Rural						
9020759	155	0	State 2			Rural						
9020595	148	0	State 2			Rural						
1021545	146	0	State 2			Rural						
10022650	145	0	State 2			Rural						

6020077	136	0	State 2			Rural						
6020812	135	0	State 2			Rural						
6022517	128	0	State 2			Rural						
3022645	124	0	State 2			Rural						
9020797	123	0	State 2			Rural						
1022680	121	0	State 2			Rural						
4022726	111	0	State 2			Rural						
2020752	105	0	State 2			Rural						
7020772	105	0	State 2			Rural						
7020594	104	0	State 2			Rural						
10022640	100	0	State 2			Rural						
7020755	98	0	State 2			Rural						
10020521	82	0	State 2			Rural						
1020574	70	0	State 2			Rural						
:	:	:	:			:						
:	:	:	:			:						

ST8B: PISA 2012 Main Survey Sampling Frame and School Exclusion File: Sheet 2 (EXAMPLE)
--

See Section 7.2.2.3 of School Sampling Preparation Manual. The Excel template for Sampling Task 8B may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/

National School ID	PISA enrolment	Reason for Exclusion
SE0005	4	Special school for students with special needs
SE0011	2	Special school for students with special needs
SE0020	3	Special school for students with special needs
:	:	:
:	:	:
Total	1162	

ST9: PISA 2012 Main Survey School Sample Selection: Sample Allocation by Explicit Strata (EXAMPLE)

See Appendix B.

INITIAL ALLOCATION

Explicit Strata labels	Population figures			Sample allocation				
	STIDSTRT	% of total ENR	Number of Schools	Number of schools to sample	Rounded	Adjusted	Estimated number of students	% of eligible students expected in sample
State 1	01	8.45%	66	12.6817057	13	13	455	8.7%
State 2	02	72.45%	1128	108.670668	109	109	3815	72.7%
State 3	03	9.01%	117	13.5123603	13	13	455	8.7%
State 4	04	10.09%	206	15.1352658	15	15	525	10.0%
Grand Total		100.00%	1517	150	150	150	5250	100.0%

Number of schools to start with: 150

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis (WORKSHEET EXAMPLE)

See Appendix A.

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **15 November 2011**

Overall or Explicit Stratum Label: **Overall** TCS: **35**

1. Enrolment in small schools:

Type of school based on enrolment	# of schools	# of students	Percentage of total enrolment
Enrolment of 15-year-old students < 18	133	2067	[a] 2.0 %
Enrolment of 15-year-old students ≥ 18 and < 35	411	10231	[b] 9.7 %
Enrolment of 15-year-old students ≥ 35	973	93161	[c] 88.3 %
TOTAL	1517	105459	100 %

2. If the percentage in box [a] is 1 percent or more and the percentage in box [b] is 4 percent or more, then very small schools will be undersampled and the school sample size will be increased.

box [a] ≥ 1% and box [b] ≥ 4%?

Yes or No



Undersample very small schools and calculate final school sample size. Done.

3. If the percentage in box [a] is 1% or more, very small schools will be undersampled and the school sample size will be increased.

box [a] ≥ 1%?

Yes or No



Undersample very small schools and calculate final school sample size. Done.

4. If the percentage in box [a] is less than 1 percent and the percentage in box [b] is 4 percent or more, then there is no undersampling of very small schools but the school sample size will be increased.

box [a] < 1% and box [b] ≥ 4%?

Yes or No



Calculate final school sample size. Done.

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis (WORKSHEET EXAMPLE)

See Appendix A. Each explicit stratum will need a separate Small School Analysis

PISA Participant: **ATLANTA**

National Project Manager: **Mr X**

Date this version of this form was completed: **15 November 2011**

Overall or Explicit Stratum Label: **State 1** TCS: 35

1. Enrolment in small schools:

Type of school based on enrolment	# of schools	# of students	Percentage of total enrolment
Enrolment of 15-year-old students < 18	0	0	[a] 0.0 %
Enrolment of 15-year-old students ≥ 18 and < 35	2	56	[b] 0.6 %
Enrolment of 15-year-old students ≥ 35	64	8860	[c] 99.4 %
TOTAL	66	8916	100 %

2. If the percentage in box [a] is 1 percent or more and the percentage in box [b] is 4 percent or more, then very small schools will be undersampled and the school sample size will be increased.

box [a] ≥ 1% and box [b] ≥ 4%?

Yes or No



Undersample very small schools and calculate final school sample size. Done.

3. If the percentage in box [a] is 1% or more, very small schools will be undersampled and the school sample size will be increased.

box [a] ≥ 1%?

Yes or No



Undersample very small schools and calculate final school sample size. Done.

4. If the percentage in box [a] is less than 1 percent and the percentage in box [b] is 4 percent or more, then there is no undersampling of very small schools but the school sample size will be increased.

box [a] < 1% and box [b] ≥ 4%?

Yes or No



Calculate final school sample size. Done.

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis (EXAMPLE)

See Appendix A. The Consortium sends a completed Small School Analysis (SSA) to the NPM as an Excel spreadsheet. The SSA columns below illustrate the analysis for any small school undersampling and/or sample size increase.

Explicit Stratum Label		P1	P2	P1+P2 ("P")	Q	R	Grand Total	% in P1	% in P2	% in P	% in Q	% in R	Is increase in sample size required?	Is undersampling Very Small Schools required?
								("very small" schools with 0,1,2)	("very small" schools with 3 or more)	("very small" schools)	("moderately small" schools)	("large" schools)		
Overall SSA	Sum of PISA ENR	0	2067	2067	10231	93161	105459	0.0%	2.0%	2.0%	9.7%	88.3%	Yes	Yes
	Count of NATID	0	133	133	411	973	1517							
State 1	Sum of PISA ENR	0	0	0	56	8860	8916	0.0%	0.0%	0.0%	0.6%	99.4%	No	No
	Count of NATID	0	0	0	2	64	66							
State 2	Sum of PISA ENR	0	1725	1725	8176	66501	76402	0.0%	2.3%	2.3%	10.7%	87.0%	Yes	Yes
	Count of NATID	0	111	111	330	687	1128							
State 3	Sum of PISA ENR	0	65	65	318	9117	9500	0.0%	0.7%	0.7%	3.3%	96.0%	No	No
	Count of NATID	0	4	4	12	101	117							
State 4	Sum of PISA ENR	0	277	277	1681	8683	10641	0.0%	2.6%	2.6%	15.8%	81.6%	Yes	Yes
	Count of NATID	0	18	18	67	121	206							

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis (EXAMPLE-Continued)

See Appendix A. The Consortium sends a completed Small School Analysis (SSA) to the NPM as an Excel spreadsheet. The columns below illustrate the calculations for sample size increase due to small schools. The base number of schools/students in individual stratum calculations follow from the sample allocation.

Explicit Stratum Label	Base number of schools	TCS	Base number of students	L (see Samp Man Apndx A)	P1 Under sampling factor	P2 Under sampling factor	V1ENR	V2ENR	MENR	Number of Very Small Schools (P1) expected to sample	Number of Very Small Schools (P2) expected to sample	Number of Moderately Small Schools expected to sample	Number of Large Schools expected to sample	Approx yield (before considering student /school non-response)	Minimum number of schools to sample
Overall SSA	150	35	5250	1.01	0.25	0.5	0	15.5414	24.8929	0	4	21	134	5275	159
State 1	13	35	455	1	1	1	0	0	0	N/A	N/A	N/A	13	455	13
State 2	109	35	3815	1.0115	0.25	0.5	0	15.5405	24.7758	0	3	17	96	3828	116
State 3	13	35	455	1	1	1	0	0	0	N/A	N/A	N/A	13	455	13
State 4	15	35	525	1.013	0.25	0.5	0	15.3889	25.0896	0	0	3	12	495	15

TOTAL 0 3 20 134 5233 157

ST9: PISA 2012 Main Survey School Sample Selection: Final Sample Allocation by Explicit Strata (EXAMPLE)

See Appendix B.

FINAL ALLOCATION

SAMPLE ALLOCATION		Population Figures					
Explicit Stratum Label	STIDSTRT	Very Small Schools (1) ENR	Very Small Schools (2) ENR	Moderately Small Schools ENR	Large Schools ENR	Grand Total ENR	Number of Schools
State 1	01	0	0	56	8860	8916	66
State 2	02	0	1725	8176	66501	76402	1128
State 3	03	0	65	318	9117	9500	117
State 4	04	0	277	1681	8683	10641	206
Grand Total		0	2067	10231	93161	105459	1517

SAMPLE ALLOCATION		Number to Sample										
Explicit Stratum Label	STIDSTRT	Number of Very Small Schools (1) expected to sample	Number of Very Small Schools (2) expected to sample	Number of Moderately Small Schools expected to sample	Number of Large Schools expected to sample	Total Number of schools	Certainty Schools Stratum	Sample drawn from original stratum (non-certainty schools)	Estimated student yield	% of eligible students expected in sample	% Total ENR	
State 1	01	0	0	0	13	13	13	0	13	455	8.7%	8.5%
State 2	02	0	3	17	96	116	116	0	116	3828	73.2%	72.4%
State 3	03	0	0	0	13	13	13	0	13	455	8.7%	9.0%
State 4	04	0	0	3	12	15	15	0	15	495	9.5%	10.1%
Grand Total		0	3	20	134	157	157	0	157	5233	100.0%	100.0%

ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata (EXAMPLE)

See Appendix B.

STIDSTRT	Cross-classification of implicit stratifiers within Explicit Stratum ID	Number of Schools	Total PISA enrolment	Total PISA Measure of Size (with adjusted small school MOS)
01		66	8916	8930
	1,1	62	8141	8155
	1,2	4	775	775
02		1128	76402	79993
	2,1	1077	70228	73751
	2,2	51	6174	6242
03		117	9500	9677
	3,1	9	658	659
	3,2	108	8842	9018
04		206	10641	11343
	4,1	206	10641	11343
Grand Total		1517	105459	109943

**ST10: PISA 2012 Main Survey School Sample Selection: Selection Numbers
(EXAMPLE)**

See Appendix B.

PISA Sampling - Selection Numbers

NPM: Mr. X

STIDSTRT: 03

Sample based on file: **ATL_Sample_WorkingFile.xls**

Date sample selected: **15/11/2011**

S	9677
(Total Measure of Size)	

D	13
(Desired Sample Size)	

I	744.3846
Sampling Interval	

RN	0.2878
Random Number	

Line Numbers		Selection Numbers
1	(744.3846 x 0.2878 =)	214.2339
2	(744.3846 x 1.2878 =)	958.6185
3	(744.3846 x 2.2878 =)	1703.0031
4	(744.3846 x 3.2878 =)	2447.3877
5	(744.3846 x 4.2878 =)	3191.7723
6	(744.3846 x 5.2878 =)	3936.1569
7	(744.3846 x 6.2878 =)	4680.5415
8	(744.3846 x 7.2878 =)	5424.9261
9	(744.3846 x 8.2878 =)	6169.3107
10	(744.3846 x 9.2878 =)	6913.6953
11	(744.3846 x 10.2878 =)	7658.0799
12	(744.3846 x 11.2878 =)	8402.4645
13	(744.3846 x 12.2878 =)	9146.8491

Appendix D - Sampling Form Templates

ST7A: PISA 2012 Main Survey National DESIRED Target Population

See Section 7.1.1 of School Sampling Preparation Manual.

Please complete this form in full before filling out ST7B. Enter information in the cells that are highlighted in green. Make sure number values have no commas or decimal points. Please complete all highlighted sections of this form. After completing this form please do ST7B.

	PISA Participant:	<input type="text"/>
	Person who completed this form:	<input type="text"/>
	Date this version of this form was completed:	<input type="text" value="d d - m m - y y y y"/>
1	Total national population of 15-year-olds:	<input type="text"/>
2.1	Total national population of 15-year-olds enrolled in educational institutions and who are in grades 7 and higher:	<input type="text"/>
2.2	Total number of 15-year-olds enrolled in grades 6 and below:	<input type="text"/>
3.1	If the national desired target population for your country differs from the total national population of enrolled 15-year-olds in grades 7 and higher, describe the population(s) to be excluded from the total national population of enrolled 15-year-olds in grades 7 and higher. These include distinct geographic regions of the country, or specific language groups. If no population is to be excluded please enter "None".	<input type="text"/>
3.2	Total enrolment <u>excluded</u> from the total national population of enrolled 15-year-olds in grades 7 and higher (corresponding to the exclusions listed above):	<input type="text"/>
4	Total enrolment in the national desired target population:	0
5	Percentage of coverage of nationally enrolled 15-year-olds, in the national desired target population:	-
6	Describe your data source and year of data. Also provide webpage address with relevant tables. Tables not in English should be emailed to Westat with English labels added for relevant table variables:	<input type="text"/>

Please include any additional comments you may have in

the box below:



You have not yet completed ST7A. Please complete all sections highlighted in green before continuing to ST7B

End of ST7A

ST7B: PISA 2012 Main Survey National DEFINED

Target Population

See Section 7.1.2 of School Sampling Preparation Manual.

Please complete ST7A before you begin to fill out this form. Enter the relevant information in the cells that are highlighted in green. Make sure number values have no commas or decimal points. Please complete all highlighted sections of this form.

PISA Participant: **Please complete ST7A**

Person who completed this form: **Please complete ST7A**

Date this version of this form was completed: **dd-mm-yyyy**

1 Total enrolment in the national desired target population: **Please complete ST7A**

2.1 School-level exclusions:

Please write a brief description of each type of school-level exclusion in the table below and for each type of exclusion you enter, please include the number of schools and students to whom this exclusion will apply.

	<u>Description of school-level exclusions</u>	<u>Number of schools</u>	<u>Number of students</u>
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
	TOTAL	0	0

2.2 Percentage of students not covered due to school-level exclusions: -

3 Total enrolment in national defined target population before within-school exclusions: -

4 Do you plan to exclude schools which have only one or two PISA students in the field during data collection and only where this will not cause an entire stratum to be missing from the collected data?

5.1 Anticipated within-school exclusions (students who could not be included in the PISA assessment, from schools where some students could be included). Please enter 0 (zero) when no students of the type will be excluded.

<u>Descriptions of within-school exclusions</u>	<u>Expected number of students</u>
Functionally disabled students	
Intellectually disabled students	
Students with insufficient language experience in test language	
Other	
No materials available in student's language of mathematics instruction	
TOTAL	0

5.2 Expected percentage of students not covered due to within-school exclusions:

-

6 Total enrolment in national defined target population:

-

7 Coverage of national desired target population:

-

8 Describe your data source and year of data. Also provide webpage address with relevant tables. Tables not in English should be emailed to Westat with English labels added for relevant table variables:

Please include any additional comments you may have in the box below:

You have not completed ST7B. Please complete the sections highlighted in green before submitting this form.

End of ST7B

ST8A: PISA 2012 Main Survey Sampling Frame Information

See Section 7.2.1 of School Sampling Preparation Manual.

Enter information in the cells that are highlighted in green. Please complete all highlighted sections of this form, unless otherwise specified. In many of the cells you will need to choose a response from a drop-down list. To do this, first select the cell, then click on the arrow symbol that appears and choose the appropriate response from the list. An error message will appear if an invalid response is selected.

PISA Participant:

Person who completed this form:

Date this version of this form was completed:

1 Will a sampling frame of geographic areas be used?

2 Specify the PSU Measure of Size to be used:

3 Specify the school year for the data which will be used for the PSU Measure of Size:

4 Please provide a preliminary description of the information available to construct the area frame. Please consult with Westat for support and advice in the construction and use of an area-level sampling frame.

5 Specify the school estimate of enrolment of 15-year-olds that will be used.

6 Specify the year for which enrolment data will be used for the school estimate of enrolment of 15-year-olds:

7 Please describe any other type of frame, if any, that will be used. If no other type of frame will be used please enter "None".

8 Please indicate whether the following types of schools exist in your system and whether they have been included or excluded in your PISA sampling frame. Each school type requires a response.

<u>School Type</u>	<u>Included / Excluded / Don't Have Any</u>
Bilingual schools	
Immersion schools	
National schools with single (minor) language of instruction	
International / foreign schools with single (minor) language	
Schools with mixed (major and minor) languages of instruction	
Schools with fewer than 3 estimated PISA students	
Non-publically funded schools	
Rural schools	
Urban schools	
Remote schools	
ISCED2 only schools	
ISCED3 only schools	
Mixed ISCED schools	
Pre-vocational schools	
Vocational schools	
Technical schools	
Agriculture schools	
Other vocational schools	
General schools	
Mixed programme schools	
Predominantly male schools	
Predominantly female schools	
All male schools	
All female schools	
Schools with only part-time students	

Schools with multiple shifts	
Religious schools	
Schools with predominantly international / foreign students	
Schools with mixed national / foreign students	
SEN only schools	
Any other irregular schools	

You have not completed question 8. Please complete question 8 before submitting this form.

Please include any additional comments you may have in the box below:

You have not yet completed ST8A. Please complete all relevant sections highlighted in green.

End of ST8A

**ST8B: PISA 2012 Main Survey Sampling Frame and School Exclusion:
Sheet 1**

See Section 7.2. The Excel template for Sampling Task 8B may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/

National School ID	PISA enrolment	Enrolment of additional population or of PISA sub-population as defined on ST1	First Explicit Stratification Variable Specified on ST2	Second Explicit Stratification Variable Specified on ST2	...	First Implicit Stratification Variable Specified on ST2	Second Implicit Stratification Variable Specified on ST2	...	Extra Variable 1 (renamed by country)	Extra Variable 2 (renamed by country)	...	Extra Variable n (renamed by country)

Note: Additional stratification variable columns (not included here due to page width) are available in the Excel template of Sampling Task 8b on MyPISA. Unshaded columns in this illustration are for required information. Additional school information as needed or desired may be added as extra variables.

**ST8B: PISA 2012 Main Survey Sampling Frame and School Exclusion:
Sheet 2**

See Section 7.2.2.3 of School Sampling Preparation Manual. The Excel template for Sampling Task 8B may be downloaded from the MyPISA website from Home/Materials/2012 MS Resources/Sampling/

National School ID	PISA enrolment	Reason for Exclusion

ST9: PISA 2012 Main Survey School Sample Selection: Sample Allocation by Explicit Strata: Initial Allocation Sheet

See Appendix B. The Consortium will do these calculations as part of the preparation for sampling and will provide this information to the NPM for sample size and sample allocation discussion (Sampling Task 9). Countries doing their own school sampling need to perform these calculations as part of the PISA school sampling process.

INITIAL ALLOCATION

Explicit Strata labels	Population figures			Sample allocation				
	STIDSTRT	% of total PISA enrolment	Number of Schools	Number of schools to sample	Rounded	Adjusted	Estimated number of students	% of eligible students expected in sample
	01							
	02							
	03							
	04							
	05							
	etc.							
	Grand Total	100.00%						100.0%

Number of schools to start with:

ST9: PISA 2012 Main Survey School Sample Selection: Small School Analysis Worksheet

See Appendix A. The Consortium will do these calculations as part of the preparation for sampling and will provide this information to the NPM for sample size and sample allocation discussion (Sampling Task 9). Countries doing their own school sampling need to perform these calculations as part of the PISA school sampling process.

PISA Participant: _____

National Project Manager: _____

Date this version of this form was completed: _____

Overall or Explicit Stratum Label: _____ TCS: _____

1. Enrolment in small schools:

Type of school based on enrolment	# of schools	# of students	Percentage of total
Enrolment of 15-year-old students < (TCS/2)			[a]
Enrolment of 15-year-old students ≥ (TCS/2) and < TCS			[b]
Enrolment of 15-year-old students ≥ TCS			[c]
TOTAL			100%

2. If the percentage in box [a] is 1 percent or more and the percentage in box [b] is 4 percent or more, then very small schools will be undersampled and the school sample size will be increased.

box [a] ≥ 1% and box [b] ≥ 4%? Yes or No
 ↓
 Undersample very small schools and calculate final school sample size. Done.

3. If the percentage in box [a] is 1% or more, very small schools will be undersampled and the school sample size will be increased.

box [a] ≥ 1%? Yes or No
 ↓
 Undersample very small schools and calculate final school sample size. Done.

4. If the percentage in box [a] is less than 1 percent and the percentage in box [b] is 4 percent or more, then there is no undersampling of very small schools but the school sample size will be increased.

box [a] < 1% and box [b] ≥ 4%? Yes or No
 ↓
 Calculate final school sample size. Done.

ST9: PISA 2012 Main Survey School Sample Selection: Sample Allocation by Explicit Strata: Final Sample Allocation Sheet

See Appendix B. The Consortium will do these calculations as part of the preparation for sampling and will provide this information to the NPM for sample size and sample allocation discussion (Sampling Task 9). Countries doing their own school sampling need to perform these calculations as part of the PISA school sampling process. The Population Figures and Number to Sample columns of the Sample Allocation Sheet are shown separately since all columns do not fit the page width.

FINAL SAMPLE ALLOCATION		Population Figures					
Explicit Stratum Label	STIDSTRT	Very Small Schools (1) ENR	Very Small Schools (2) ENR	Moderately Small Schools ENR	Large Schools ENR	Grand Total ENR	Number of Schools
	01						
	02						
	etc.						
	Grand Total						

FINAL SAMPLE ALLOCATION		Number to Sample										
Explicit Stratum Label	STIDSTRT	Number of Very Small Schools (1) expected to sample	Number of Very Small Schools (2) expected to sample	Number of Moderately Small Schools expected to sample	Number of Large Schools expected to sample	Total Number of schools	Certainty Adjusted	Certainty Schools Stratum	Sample drawn from original stratum (non-certainty schools)	Estimated student yield	% of eligible students expected in sample	% Total ENR
	01											
	02											
	etc.											
	Grand Total										100.0%	100.0%

ST10: PISA 2012 Main Survey School Sample Selection: Population Counts by Strata

See Appendix B. The Consortium will complete the Population Counts by Strata and return this to the NPM as a separate sheet in the frame file with sampled schools attached (Sampling Task 10). Countries doing their own school sampling will need to perform these calculations as part of the PISA school sampling process.

Explicit Stratum ID	Cross-classification of implicit stratifiers within Explicit Stratum ID	Number of Schools	Total PISA enrolment	Total PISA Measure of Size (with adjusted small school MOS)
01				
02				
etc.				
Grand Total				

Appendix E – Financial Literacy and Computer-based Assessments

E.1 Financial Literacy

177. Financial Literacy (FL) is a paper-based International Option for PISA 2012. Financial Literacy (FL) is a paper-based International Option for PISA 2012. There will be four FL paper-based booklets for the MS. In each booklet there will be two clusters of FL material, one cluster of mathematics material and one cluster of reading material so that the testing time for FL students will be the same 2 hours as for the PISA paper-based assessments (PBAs).

E.1.1 Student Sample Size

178. As in the FT, the FL booklets will be rotated over the sampled students. The sample size of FL students for the MS is at least 1200 students. This means that with a usual minimum 5250 students required to be sampled for PISA PBAs, the total minimum sample size for all PBAs is $5250 + 1200 = 6450$.

179. It is assumed that if a country is participating in the FL PBA option, then it is expected that FL assessments will occur in every PISA sampled and participating school. This then means that if a base school sample size of 150 schools was assumed, the number of students to sample in each school would be 43, up from the usual 35. Of the 43 sampled students in each school, about 8 would be assigned the FL PBA and the other 35 sampled students would get the PISA PBA.

180. If generally the target cluster size of students to sample per school for a country in previous PISA cycles has been less than 35, say 25, for example, then targeting 5250 students means a larger school sample size (of 210 schools in this example). Then the target of $5250 + 1200 = 6450$ students overall would mean a target cluster size of 31 students per school in this case. This would mean on average 6 students per school would get paper-based FL and 25 students would get paper-based PISA.

181. Consider some other scenarios. If generally the target cluster size of students to sample for PISA is usually, say 40, then since no fewer than 150 schools can be sampled for PISA, that means the target is 6000 PISA paper-based students. If such a country is doing FL, then the total target number of students to sample is $6000 + 1200 = 7200$. This means that 48 students should be sampled per school. However, since the rotation of FL booklets is in the ratio of 8/43, then $8/43 \times 48$ is about nine students for FL, with 39 having paper-based PISA. If it was important to the

country to have 40 paper-based PISA students per school, then the target cluster size per school should be increased from 48 to 49. On the other hand, if it is important to the country to not have any more than 40 students sampled per school for PBAs, then if the target cluster size was left at 40, this would mean on average seven FL students per school and only 33 for paper-based PISA. With 150 sampled schools, 33 paper based PISA students would not meet the requirements of at least 5250 sampled students for paper based PISA. Therefore, the best that can be done in this case is either to sample 150 schools and 43 students per school, or to increase the number of sampled schools to 165 and maintain a target cluster size of 40. Both satisfy the minimum requirements of having 5250 students sampled for paper based PISA and 1200 for paper based FL.

182. Note that in the last example, if the target cluster size had been 40 because of student response rate concerns, then it is likely best that the number of FL students to target also be raised to about 1400.

E.1.2 School Sample Size

183. As just indicated, the target cluster size of students to sample for PBAs in each school is larger than usual when FL is being done. Recall that a school is called a small school if it has fewer PISA students than the target cluster size. Countries doing FL will have more small schools than if they were not doing FL. It is thus likely that because of the small school analysis done in each explicit stratum, the required school sample size will be larger than in the previous cycle, all else equal.

E.1.3 Non-Participation

184. Students assigned to do a FL PBA are PISA students just like those assigned to do a PISA PBA. Therefore any sampled student who refuses to participate, no matter which type of paper-based assessment he/she is assigned, should be given the appropriate student non-participation code as usual.

E.1.4 SEN students

185. It is expected that PISA students in SEN schools will participate in the FL option. Countries should indicate to Westat if their plans are contrary to this, even if this has previously been discussed, as circumstances can change. If the country uses the UH booklet in SEN schools, a special Financial Literacy UH (FLUH) booklet has been created and will be rotated through the sampled students at a rate of 8/43.

186. Similarly, it is expected that SEN PISA students in mainstream schools who would have been excluded but will instead use the UH booklet, will also participate in the FL option. FL booklets (FL or FLUH) are rotated over all sampled students at a rate of 8/43. Countries should indicate to Westat if their plans are contrary to this, even if this has previously been discussed, as circumstances can change.

E.2 Computer-based Assessments

187. For PISA 2012, computer- based assessments (CBAs) can consist of general Problem Solving (PS) only, or of Problem Solving and Computer-Based Assessment of Literacies (Mathematics and Reading) (CBAL) together. PS is a core computer based component while CBAL is a computer-based International Option.

188. In the FT, domain specific forms were rotated over subsampled students. This will not be the case for the MS. Instead, MS forms can have mixed domain clusters. If a country is doing PS only, then the CBA forms will have only PS clusters. However, if a country will do PS and CBAL, then the computer based forms will have a mixture of PS clusters, computer-based mathematics clusters and computer-based reading clusters.

E.2.1 Student Sample Size

189. If a country is doing PS only, then at least 2100 PS students need to be sampled. These PS students will be subsampled from the students sampled for paper-based PISA. If we assume 150 schools will be sampled to obtain at least 5250 sampled students for paper based PISA, then this means that at least 14 students should be subsampled for PS from the PISA sampled students in each school.

190. If generally the target cluster size of students to sample per school for a country in previous PISA cycles has been less than 35, say 25, for example, then targeting 5250 students means a school sample size of 210 schools. In this case, then at least 10 students should be subsampled for PS from the PISA sampled students in each school.

191. If a country is doing PS and CBAL, then at least 2700 students need to be sampled for CBAs. These CBA students will be subsampled from the students sampled for paper-based PISA. If we assume 150 schools will be sampled to obtain at least 5250 sampled students for paper-based

PISA, then this means that at least 18 students should be subsampled for CBAs from the PISA sampled students in each school.

192. If generally the target cluster size of students to sample per school for a country in previous PISA cycles has been, say 25, for example, then targeting 5250 students means a school sample size of 210 schools. In this case, then at least 13 students should be subsampled for CBAs from the PISA sampled students in each school.

193. If a country which is doing CBAs is also doing the FL PBA, students who are assigned the Financial Literacy assessment are **not** also assigned the usual PISA PBA. FL students are not eligible for a CBA and will not be sampled for CBA.

194. If a sampled school does not have enough working computers for the number of students who should do the CBA, arrangements will need to be made to either bring in laptops, or to have extra sessions using the available working computers. It is important to accommodate the minimum number of students to be subsampled for CBA within each school.

195. Any small sampled PISA schools, even those with only one, or two, or three PISA students (as examples) still need to have the CBA done. In such cases, all PISA PBA students will be doing PISA PBA and CBA, but any students sampled for FL, if a country is doing FL, will not do CBA.

E.2.2 School Sample Size

196. Since students for the CBA are subsampled from PISA (non-FL) paper-based sampled students, if a school sample meets the paper-based sample size requirements, no increase in school sample size is needed for the CBAs.

197. If a country is participating in any CBAs, then it is expected that student subsampling for CBA will occur in every PISA sampled and participating school to select students for the CBAs. Countries should indicate to Westat if their plans are contrary to this, even if this has previously been discussed, as circumstances can change. In particular, if a country has a large PISA school sample and wishes to subsample the PISA sampled schools where student subsampling for CBAs will be done, this becomes an additional national option, with the associated fees.

E 2.3 Non-Participation

198. If a sampled school for PISA agrees to do PISA but not to do the computer based component, such a school should **not** be replaced. The PISA school will have a school participation status in

KeyQuest indicating participation as long as at least one sampled non-excluded PISA student is assessed. The students subsampled for the CBA will all be noted as absent in this case.

199. If a sampled school agrees to do PISA but not to do the CBAs, this should be noted in the COMM space of the SCP Instrument in *KeyQuest*. See the DMM for further details.
200. There is no longer a restriction that a CBA student's data will not be used unless they have also done the PBA. However, note that if the within school scalability rate based on PISAPB and CBA, is below 25%, no PBA nor CBA data will be used (since such a school is noted in weighting as a non-respondent and all student data is dropped).
201. No non-response adjustments will be made for schools or students sampled for CBAs which do not participate. Since CBA domains are being treated as minor domains, like for example science, students assigned to the CBA assessments will be treated in a somewhat similar manner as a student assigned a booklet containing items in the science domain. Plausible values (PV) will be generated for students who were sub-sampled to do a CBA and are deemed as having PISA scalable data, as well as for those not sub-sampled for CBA and are deemed as having PISA scalable data. The scalable PISA data can be either PISA data, CBA data, or both. The only limiting condition is that there are enough CBA participants in each school on which to base the PV generation. The minimum number of CBA participating students for reliable CBA PV generation in each school is being investigated by ACER. For now, please note that as long as 70% or more of the CBA sampled students within a school participate in CBA assessment, the school is considered as participant for the purpose of determining CBA response rates and for reliable PV generation. If that condition is not met, students sampled and eligible for CBA in such a school may be treated as CBA non-respondents, even though some CBA students may have participated. **The 70% within school participation condition should be considered as a guideline and will be refined after investigation of the minimum number of CBA participating students for reliable CBA PV generation is completed.**
202. Because PS is a core domain, the requirement of an 80% student response rate (nationally) also applies for CBA.
203. Some examples follow to illustrate these points. Note that in School A, the minimum number of CBA students have not been assessed, so therefore the CBA student response rate in this school is 0.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School A		Yes (97%)	No		0.67		
# PB sampled	35						
#PB eligible& not excluded	34	denominator: 34		Denominator: 34			
# PB scalable	30			Numerator: 30			
# CBA sampled	18				Denominator: 18		
#CBA eligible& not excluded	17						
# computers available	16					Denominator: 16	
# CBA scalable	12				Numerator: 12	Numerator: 0	
# PB scalable or CBA scalable	33	numerator: 33					33

204. For school B we see that since at least 70% of CBA sampled students have been assessed, the CBA student response rate is not zero, and instead is calculated with the denominator equal to the minimum of (# of available computers, # CBA eligible and not excluded). When not enough computers are available, because CBA students are dropped randomly, these cannot cause any nonresponse bias and so do not factor into the CBA student response rate.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School B		Yes (88%)	No		0.72		
# PB sampled	35						
#PB eligible& not excluded	34	denominator: 34		Denominator: 34			
# PB scalable	30			Numerator: 30			
# CBA sampled	18				Denominator: 18		
#CBA eligible& not excluded	17						
# computers available	15					Denominator: 15	

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
# CBA scalable	13				Numerator: 13	Numerator: 13	
# PB scalable or CBA scalable	30	numerator: 30					30

205. School C is a case that in PISA 2009 would have been classed as a PISA nonrespondent by the < 25% scalable rule and the data would have been discarded, with no affect on student response rates. However, for 2012, since the scalability of all students is considered, this school is considered as a respondent for response rates. As a result, this poor PISA PB response rate will now be part of the student response rate calculation. For CBA, not enough students were assessed, so the CBA response rate in this school is zero.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School C		Yes (51%)	No		0.56		
# PB sampled	35						
#PB eligible& not excluded	34	denominator: 34		Denominator: 34			
# PB scalable	8			Numerator: 8			
# CBA sampled	18				Denominator: 18		
#CBA eligible& not excluded	14					Denominator: 14	
# computers available	15						
# CBA scalable	10				Numerator: 10	Numerator: 0	
# PB scalable or CBA scalable	18	numerator: 18					18

206. School D is similar to school C but the difference is that the CBA student response rate in school D is 100%.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School D		Yes (56%)	No		0.77		
# PB sampled	35						
#PB eligible& not excluded	34	denominator: 34		Denominator: 34			
# PB scalable	8			Numerator: 8			
# CBA sampled	18				Denominator: 18		
#CBA eligible& not excluded	14					Denominator: 14	
# computers available	15						
# CBA scalable	14				Numerator: 14	Numerator: 14	
# PB scalable or CBA scalable	19	numerator: 19					19

207. School E shows that this school, based on the scalability of both PISA PB and CBA, will be considered a nonrespondent for response rates, but the students will be retained on the file. A consequence of the < 50% rule is that these students will not be used at all in student response rate calculation, for PISA PB nor CBA, despite the fact that there were enough CBA students assessed and the CBA student response rate would have been very good. Since these students will be retained on the file, all 17 which had some data will have scores for all domains. Perhaps note that, because of the poor PBA response rate, there is not a good basis for imputing scores in ANY of the domains for many of the 35 students, and this is why it is considered a school nonrespondent, even though the CBA student response rate was high.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School E		No (49%)	No		0.94		
# PB sampled	35						

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
#PB eligible¬ excluded	35	denominator: 35		Denominator: 0			
# PB scalable	16			Numerator: 0			
# CBA sampled	18				Denominator: 18		
#CBA eligible¬ excluded	18					Denominator: 0	
# computers available	18						
# CBA scalable	17				Numerator: 17	Numerator: 0	
# PB scalable or CBA scalable	17	numerator: 17					17

208. School F is similar to School E in its outcome. The only difference is that School F did not have enough CBA assessed students so its CBA student response rate would have been zero. Instead, none of these students will be used for student response rates. Since the students in this school will be retained on the file, all 10 which had some data will have scores for all domains.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School F		No (40%)	No		0.56		
# PB sampled	35						
#PB eligible¬ excluded	25	denominator: 25		Denominator: 0			
# PB scalable	0			Numerator: 0			
# CBA sampled	18				Denominator: 18		
#CBA eligible¬ excluded	13					Denominator: 0	
# CBA scalable	10				Numerator: 10	Numerator: 0	
# computers available	18						
# PB scalable or CBA scalable	10	numerator: 10					10

209. School G doesn't have enough scalable students. Since fewer than 25% of the students are scalable, this school is a nonrespondent and all of its students are discarded.

		school considered PISA respondent by >50% scalable rule	school considered PISA nonrespondent by <25% scalable rule (student data not kept)	PB student response rate contribution	Proportion of required CBA assessments	CBA student response rate contribution	Students receiving scale scores
School G		No (23%)	Yes		0.22		
# PB sampled	35						
#PB eligible¬ excluded	35	denominator: 35		Denominator: 0			
# PB scalable	4			Numerator: 0			
# CBA sampled	18				Denominator: 18		
#CBA eligible¬ excluded	18					Denominator: 0	
# CBA scalable	4				Numerator: 4	Numerator: 0	
# computers available	10						
# PB scalable or CBA scalable	8	numerator: 8					0

E 2.4 SEN Students

210. It is expected that SEN PISA students in SEN schools will participate in any CBAs that the country is implementing in other non-SEN schools. Countries should indicate to Westat if their plans are contrary to this, even if this has previously been discussed, as circumstances can change.

211. Similarly, it is expected that SEN PISA students in mainstream schools who would have been excluded but will instead use the UH booklet, will also participate in CBAs. Countries should indicate to Westat if their plans are contrary to this, even if this has previously been discussed, as circumstances can change. Note, however, that there is no UH component for any of the CBA assessments. SEN students who require a UH booklet and are subsampled for CBA will be expected to take a one hour paper-based assessment with a PISA UH booklet, followed later by a 40-minute CBA assessment.

Appendix F – The Link to PISA 2003 International Option

F.1 Overview

212. A proposal accepted for PISA 2012, under Module E, is for a school-level longitudinal component. For countries interested in this component, which participated in 2003, the 2012 school sample would be designed to have maximum overlap with the 2003 sample, while remaining fully representative of the 2012 population. School Questionnaire data, and student achievement data summarized at the school level, for both 2003 and 2012 could be incorporated into a school level data set for analysis. Alternatively, school-level data from 2003 could (along with 2012 school-level data) be used for student-level analysis from 2012. This might present an opportunity to examine whether changes in school policies and practices over a nine-year period are related to student achievement.
213. This procedure can only be done if the same national school identifiers are used on the PISA 2003 and PISA 2012 school frames for the schools that are the same on each frame. This responsibility falls on the NPM to ensure that the national identifiers on the 2012 frame are the same as those on the 2003 frame.
214. In what follows, we are concerned only with 2003 frame schools that are still on the 2012 frame. Any 2003 schools no longer on the 2012 frame could not be sampled for PISA 2012 and therefore could not form part of the link sample.

F.2 The Overlap Control Method

215. This maximizing of the overlap between PISA 2012 sampled schools and PISA 2003 sampled schools can occur under several different scenarios, all involving a comparison of the 2003 probabilities of selection with the 2012 probabilities of selection for the same schools on both frames. Scenario A has ALL PISA 2012 probabilities of selection *larger than or equal to* the school's PISA 2003 probabilities of selection. Scenario B has ALL PISA 2012 probabilities of selection *smaller than or equal to* the 2003 probabilities of selection. Scenario C has NOT all PISA 2012 probabilities larger than or equal to the PISA 2003 probabilities--- in other words, a mix
216. The first step before deciding which scenario applies is to link the 2003 PISA frame with the 2012 PISA frame by the national school identifier. For all schools from the 2003 frame that link to the 2012 frame, the 2003 school probability of selection and an indicator about whether or not the school was selected for 2003, are added to the 2012 frame file by ACER and Westat.

217. The next step is to do the explicit strata small school analyses for 2012 to determine explicit strata sample sizes. If the NPM accepts the overall sample size, then the PISA 2012 school probability of selection can be calculated for all schools on the 2012 frame.

218. For each school which was also on the 2003 frame, the 2012 probability of selection is compared to the 2003 probability of selection to determine which of scenarios A, B, or C applies.

F.2.A Scenario A

219. If we have scenario A, so that ALL PISA 2012 probabilities of selection are larger than or equal to those for 2003, then we apply maximum overlap control procedures in the following way.

220. To control overlap, the sample selection of schools for PISA adopts a modification of the approach due to Keyfitz (1951), based on Bayes Theorem.

221. Assume that PROBI is the PISA 2003 probability of selection, and PROBP is the PISA 2012 probability of selection (where PROBP=MOS/stratum sampling interval), then a conditional probability of selection into PISA 2012, CPROB, is determined to maximize the overlap with PISA 2003, as follows.

222.

New for 2012

$$CPROB = \begin{cases} \min \left[1, \left(\frac{PROBP}{PROBI} \right) \right] & \text{if the school was a PISA 2003 school} \\ \max \left[0, \frac{(PROBP - PROBI)}{(1 - PROBI)} \right] & \text{if the school was not a PISA 2003 school} \\ PROBP & \text{if the school was not a PISA 2003 eligible school} \end{cases}$$

223. Then a conditional MOS variable is created to coincide with these conditional probabilities as follows:

$$CMOS = CPROB \times \text{stratum sampling interval (rounded to 4 decimal places).}$$

224. The PISA 2012 school sample is then selected using the line numbers created as usual (see B.6.3), but applied to the cumulated CMOS values (as opposed to the cumulated MOS values) (see B.3.6). Note that it is possible that the resulting PISA 2012 sample size could be a bit lower or higher than the originally assigned sample size, but this is deemed acceptable.

225. Since in scenario A, ALL PROBP are greater than or equal to PROBI, then the conditional probabilities of selection for the schools also sampled in 2003 will be 1, so these schools will all be selected into the 2012 sample.

226. The PISA 2012 school weights will be calculated as usual for the 2012 sample. Since we have ensured that all schools which were sampled in 2003 and were on the 2012 frame, were also sampled for 2012, then the 2003 school weights can be used for the link sample.

F.2.B Scenario B

227. If we have scenario B, where all PISA 2012 probabilities of selection are smaller than or equal those for 2003, and apply the procedures in paragraphs 214 - 217, this will result in a sample for 2012 in which some but not all of the 2003 sampled schools are sampled for 2012. Additionally, schools on the 2003 frame which were not sampled for PISA 2003, will in this scenario of probabilities, not be sampled for the 2012 sample. Therefore, all the other schools sampled for 2012 will be those not on the 2003 frame. In this particular situation, the PISA 2012 sample can be thought of as a subsample of the 2003 sample plus a sample of *new* schools.

228. The PISA 2012 school weights will be calculated as usual for the 2012 sample. Since not all 2003 sampled schools are in the link sample, the 2003 weights cannot be used for the link sample.

229. However, the 2012 frame can be thought to consist of two parts: the 2003 frame and new 2012 schools. The 2012 sample can also be thought to consist of two parts: 1) a sample of 2003 sampled schools, and 2) a sample of “new” 2012 schools. The sum of 2012 weights for set 1) schools represents all still existing schools on the 2003 frame. The sum of 2012 weights for set 2) schools represents all the new schools on the 2012 frame which were not on the 2003 frame. Therefore, for the common schools in the link sample, these 2012 weights can also be used.

F.2.C Scenario C

230. If we have scenario C, where some PISA 2012 probabilities are larger than or equal those for 2003 and some are smaller than they were for 2003, then the following additional procedure will apply for each school on the 2012 frame. If the 2012 probability of selection is greater than or equal to the 2003 probability of selection, then PROBP remains as PROBP. If the 2012 probability of selection is less than the 2003 probability of selection, then PROBP is assigned the 2003 probability of selection.

231. Since probabilities have been adjusted, these need to be summed within explicit strata to determine revised sample sizes. If the overall sample size which now includes all 2003 schools in the 2012 sample is larger than that already discussed with the NPM after the explicit strata small school analyses, then the NPM needs to make a decision.
232. If then NPM decides that the larger sample is fine, then the 2012 sample will be selected using paragraphs 214-217. Since all PROBP will then be greater than or equal to PROBI, the maximal overlap procedure will ensure that all 2003 schools will be in the 2012 sample. The 2012 sample will use weights calculated in the usual way for the 2012 sample. The link sample will use the 2003 weights since all 2003 schools are included in the link sample.
233. If the NPM decides that the larger sample is not accepted, then we need to revert to paragraphs 214-217, using unadjusted PROBP values and the PROBI values. This will maximize overlap but it will not ensure that all 2003 sampled schools are in the 2012 sample. Nor can it ensure that non-link 2012 sampled schools are restricted to new 2012 schools.
234. The PISA 2012 weights will be calculated as usual for the 2012 sample. For the link sample, since not all 2003 sampled schools are included, the 2003 weights cannot be used to weight the link sample. Since some 2012 sampled schools may be those not sampled in 2003, then the 2012 weights cannot be used for the link sample. Thus, a special set of weights will be needed for the link sample schools, based on the probability of each school being in both the 2003 sample and the 2012 sample. Each school in the link sample would be weighted as $1 / (\text{PROBP} * \text{PROBI})$. Then school nonresponse adjustments would need to be done to account for schools that did not participate in either/or PISA 2003 and PISA 2012. This extra procedure is ONLY required when the NPM cannot accept the larger sample size noted in paragraph 224.

Appendix G – Participation in Grade or other National Option Sampling and Sample Size

G.1 Samples of Students in a Particular Grade

235. The proposal put forth by the Consortium for PISA 2012 responded to a discussion in the Call for Tender about a grade-based sampling option. The grade identified for each country should generally be the one with the greatest proportion of PISA-eligible students. In countries where two grades each have more than 40 percent of these students, the NPM can indicate which of the two grades they wish to include. In such cases, if a country elects to include both grades, the second grade becomes a second national option. In fact, any country can elect to add an additional grade of its own choosing, as a national option.

236. If direct student sampling is used, the results from this grade sample will be of comparable precision to the PISA age-based results, with perhaps some increase in precision as the population is somewhat more homogeneous in some countries, if about the same sample size of schools and grade students is assumed as that used for PISA students.

237. For each country considering this type of an option, the increase in the sample size depends upon whether or not direct student sampling is used. If direct student sampling is used, the increase in sample size then depends on the overlap between PISA-eligible students and the grade students as well as on whether more students will be selected per school, and/or more schools selected (see paragraphs 238 -246).

238. To distinguish between the sampling of PISA-eligible students and the grade-based students, the term grade-based is sometimes used. Note that some students may be both PISA-eligible and grade-eligible. However, grade-eligible students are not necessarily PISA-eligible. For example, if the grade-based option relates to students in grade 10, all grade 10 students are grade-eligible, but only those grade 10 students that fall within the ages defined in the PISA Target Population are PISA-eligible.

239. With the direct student sample, the sample of grade-eligible students would be sampled directly from a list of all such eligible students within the school. This sample selection could, and most likely would, be fully integrated with the selection of the PISA-eligible sample. Thus, participating schools would be asked to provide a student list containing all PISA-eligible students, plus all other students enrolled in grade 10, for example. A sample would be selected from this combined student list, where the total sample size would be determined in part by the

extent of overlap between the PISA-eligible and grade-eligible populations. The two samples, the PISA-eligible sample and the grade-eligible sample, will be weighted together. For analysis purposes, the students can be separated into PISA students or grade students as desired. Students who are eligible for both PISA and the grade sample can be used in both analyses, and will have the same initial student base weights in each.

240. Although direct student sampling is generally preferable, a classroom-level option is another way for the grade option to be achieved. In the case of classroom sampling, the selection of one or more classrooms from within the selected schools would be appropriate. Thus, the sample selection within a school would consist of two components. First a sample of PISA-eligible students would be selected from a list of PISA-eligible students enrolled in school. Next, a sample of grade 10 (for example) mathematics classes would be selected from a list of grade 10 mathematics classes that covered all grade 10 students in the school. All students from the selected classes would be included in the sample. Some students, who are both PISA-eligible and grade-eligible, would no doubt be selected twice. Such students would not need to be assessed twice, but the data they provide from a single assessment could be used in analyses of both the grade- and age- eligible populations, with different student weights (because of different sampling methods) for each type of analysis.

241. The classroom-based approach will present challenges in many cases, in the preparation of a suitable list of classrooms from which to select the sample. Also the best approach will depend upon whether classroom-based sampling is being used purely for operational convenience, as the easiest way to administer the assessment to a sample of grade 10 students (although this is unlikely to be the case) or whether the aim is to gather information about classrooms of students, either in association with a teacher questionnaire, or just as a means to provide good data for use in variance decomposition and mixed effects linear models (i.e., HLM).

242. Countries interested in having a national grade option are required to alert Westat to this fact when discussing the school sample. **This should not be a new option at the time of the MS as any national options, including grade sampling, were required to be field trialled.**

243. Countries wishing to have a grade sample in a sample of schools that is different than the PISA sample of schools should handle this independently of PISA, although such a national option is still subject to the Consortium approval as described below.

244. There are two ways to accomplish a direct-student sampled grade sample in KeyQuest. One option assumes that the sampling interval calculated for just the PISA-eligible students is then

applied to the full student list containing both PISA students and non-PISA grade students. The total sample size within schools have the grade of interest is variable and depends on the extent of overlap between the PISA students and the grade students. This is called WSMOP=1 in KeyQuest.

245. The second direct-student sampled grade sample is called WSMOP=2 in KeyQuest. In this sampling option, the sampling interval is calculated over the full student list and is not variable by school (unless the school is a small PISA school). This option requires additional schools to be sampled for PISA since not all sampled students in this sampling option will be PISA-eligible students.

246. For both of these sampling options to function properly in KeyQuest, the student lists need to have three sets of students identified using two student list variables, TFPOPELIG0, and TFPOPELIG1. TFPOPELIG0 needs to take value 1 for all students on the list which are PISA-eligible, and 0 otherwise. TFPOPELIG1 needs to take value 1 for all students on the list which are grade-eligible, and 0 otherwise. Students which have TFPOPELIG0=1 and TFPOPELIG1=0 are students eligible for PISA only. Students which have TFPOPELIG0=0 and TFPOPELIG1=1 are students which are grade-eligible only. Students which have TFPOPELIG0=1 and TFPOPELIG1=1 are both PISA-eligible and grade-eligible students.

247. Consider a school with 175 PISA-eligible students, 30 of which are also grade-eligible students. This school has an additional 50 grade-only students. In this school then, there would be 145 students with TFPOPELIG0=1 and TFPOPELIG1=0. There would be 30 students with TFPOPELIG0=1 and TFPOPELIG1=1. There would also be 50 students with TFPOPELIG0=0 and TFPOPELIG1=1.

248. Assume that the TCS=35. With WSMOP=1, the sampling interval is calculated as $175 / 35 = 5$. This means that out of the 145 students with TFPOPELIG0=1 and TFPOPELIG1=0, there would be 29 sampled students. Out of the TFPOPELIG0=1 and TFPOPELIG1=1 students (30), there would be 6 sampled students. Out of 50 students with TFPOPELIG0=0 and TFPOPELIG1=1, there would be 10 students sampled. There would be a total of 45 students sampled from this school using WSMOP=1. There are 35 sampled students who are PISA-eligible. There are 16 students who are grade-eligible.

249. With WSMOP=2, the sampling interval is calculated as $225 / 35 = 6.4286$. This means that out of the 145 students with TFPOPELIG0=1 and TFPOPELIG1=0, there would be 22 or 23 sampled students. Out of TFPOPELIG0=1 and TFPOPELIG1=1 students (30), there would be 4 or 5

sampled students. Out of 50 students with TFPOPELIG0=0 and TFPOPELIG1=1, there would be 7 or 8 students sampled. There would be a total of 35 students sampled from this school using WSMOP=2. There would be between 26 and 28 sampled students who are PISA-eligible. There would be between 11 and 13 students who are grade-eligible.

250. Note that with WSMOP=1, the total student sample size in a school increases as the number of grade-only students increases from the number of PISA students. For example, if the school in question had 175 PISA-eligible students, 30 of which are also grade-eligible students and 200 grade-only students, then the total student sample size would be 75 students.

251. If class sampling is desired rather than direct student sampling, then the sampling option to use in KeyQuest is WSMOP=3. For this option, it is assumed that the classes are sampled in a school first in KeyQuest before the PISA sample is selected in the school. The school's student list is then defined with TFPOPELIG0 and TFPOPELIG1 values as follows. If a student is PISA-eligible only, then TFPOPELIG0=1 and TFPOPELIG1=0. If a student is PISA-eligible and also in the sampled class(es), then TFPOPELIG0=1 and TFPOPELIG1=1. If a student is not eligible for PISA but in the sampled class(es), then TFPOPELIG0=0 and TFPOPELIG1=1. If the TCS=35 and the school has at least 35 PISA-eligible students, then KeyQuest will sample 35 PISA students from the students having either (TFPOPELIG0=1 and TFPOPELIG1=0) or (TFPOPELIG0=1 and TFPOPELIG1=1). KeyQuest will then additionally take into the school's student sample, all students with TFPOPELIG0=1 and TFPOPELIG1=1 which were not sampled in the first sample selection of 35 PISA students, as well as all students with TFPOPELIG0=0 and TFPOPELIG1=1.

252. There are additional WSMOPs available in KeyQuest for grade sampling but these are only used when suggested to a country by the Consortium after hearing about particular country circumstances, or if they have also been used in PISA 2009. They are therefore not discussed here since only used in rare cases.

253. See the DMM for further specifics about these WSMOP=1, 2, or 3 cases.

254. Countries having a class based grade sample need to include extra information on student lists. The variable, TFCLName, should contain the class name. After class sampling in KeyQuest, the variable, TFCLID, will have a class ID, and this variable also needs to be on the list of students.

255. Note that for the MS, only one WSMOP value can be used in an explicit stratum. This restriction may require the stratification plan on ST2 to be revised before MS school sampling.

256. The WSMOP assigned to original sampled schools will also be assigned to their assigned replacements.

G.2 Other National Options

257. All other national options requiring the sampling of students other than PISA or grade-eligible students are required to be approved by the Consortium and to have been field trailed, before they can be used in the MS.

Appendix H – Area-level Sampling Frames

258. Sampling units on this type of sampling frame are usually geographical areas (GAs), for which reliable information on PISA student enrolment is known. The GAs are usually identifiable geographical units used during the taking of a national census, or administrative units for which basic education statistics are available.

259. The GAs become the primary sampling units (PSUs) at this first sampling stage. The area-level sampling frame then, consists of a list of all eligible PSUs. 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 as representative of all PSUs. The PISA guidelines on this matter require a minimum of 80 PSUs in the area-level sampling frame and a minimum of 40 PSUs sampled from that frame using PPS sampling with the size ideally being the number of 15-year-olds. (Exceptions regarding these minimums should be discussed with Westat.) NPMs will need to discuss the implementation of these guidelines with Westat, 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 representative of all PSUs. The result would be a potentially low national coverage, which would be reported, as such in the analytical reports.

260. If you choose to use an area-level sampling frame, each PSU in the list should include at minimum:

- Unique PSU identification information;
- Appropriate entries for any suitable stratification variables (see ST2); and
- A suitable PSU measure of size.

261. Suitable PSU measures of size could be:

- 15-year-old student enrolment;
- Total student enrolment;
- Number of schools; or
- Population size.

262. The second level of this type of sampling frame consists of a comprehensive list of schools from all sampled PSUs. For each PSU selected from the area-level sampling frame, a comprehensive list of schools is prepared according to the guidelines in section 7.2.2.1.

263. **The school frame containing all schools that could contain PISA students at the time of the assessment for all sampled PSUs, needs to be submitted to MyPISA as Sampling Task 8B form and as one Excel file, with at least two data sheets.**

Appendix I - What is a School?

264. As noted earlier, we generally wish to sample whole schools as the first stage unit of selection, rather than programmes or tracks within schools, or shifts within schools, so that the meaning of “between school variance” is more comparable across countries.
265. There are exceptions to this, such as when school shifts are actually more like separate schools than part of the same overall school. However, in some countries with school shifts this is not the case and therefore whole schools are used as the primary sampling unit. Similarly, many countries have schools with different tracks/programs but generally we recommend again that the school as a whole should be used as the primary sampling unit. There are some exceptions, such as the schools being split for sampling in previous PISA cycles (trends would be affected if the same practice was not continued), or if there is a good reason for doing so (such as to improve previously poor response rates, differential sampling of certain tracks or programs is desired, etc).

Appendix J - Within-School Exclusion Weighted Estimates for PISA 2009 Countries

	Weighted (SEN code=1)	Weighted (SEN code=2)	Weighted (SEN code=3)	Weighted (SEN code=4)	Weighted (SEN code=5)
Albania	0.00	0.00	0.00	0.00	0.00
Argentina	287.98	937.39	0.00	0.00	0.00
Australia	272.06	2833.79	1282.75	0.00	0.00
Austria	0.00	316.51	290.12	0.00	0.00
Azerbaijan	0.00	0.00	0.00	0.00	0.00
Belgium	26.06	170.84	94.61	0.00	0.00
Belgium (Flanders)	25.06	151.00	0.00	0.00	0.00
Brazil	2495.22	196.93	0.00	0.00	0.00
Bulgaria	0.00	0.00	0.00	0.00	0.00
Canada	428.39	19082.28	1326.05	0.00	0.00
Chile	176.51	443.13	0.00	0.00	0.00
China - SAR Hongkong	0.00	118.74	0.00	0.00	0.00
China - SAR Macao	0.00	0.00	0.00	0.00	0.00
China (Shanghai)	19.26	110.90	0.00	0.00	0.00
Chinese Taipei	684.40	977.14	0.00	0.00	0.00
Colombia	200.35	47.67	242.47	0.00	0.00
Croatia	34.05	239.04	0.00	0.00	0.00
Czech Republic	116.89	143.93	161.94	0.00	0.00
Denmark	164.59	1431.71	195.50	656.11	0.00
Estonia	8.23	86.68	2.26	0.00	0.00
Finland	37.82	446.54	110.26	98.92	23.08
France	303.95	0.00	0.00	0.00	0.00
Germany	863.61	2442.69	284.65	0.00	0.00
Greece	171.89	352.28	195.06	2257.31	0.00
Hungary	0.00	48.23	0.00	313.21	0.00
Iceland	3.00	78.45	64.56	38.52	1.00
Indonesia	0.00	0.00	0.00	0.00	0.00
Ireland	51.12	783.46	261.51	395.84	0.00
Israel	194.22	1048.87	115.51	0.00	0.00
Italy	748.20	6240.94	3673.63	0.00	0.00
Japan	0.00	0.00	0.00	0.00	0.00
Jordan	166.44	148.95	127.49	0.00	0.00
Kazakhstan	428.84	827.93	0.00	0.00	2586.85
Kyrgyzstan	1093.26	211.23	79.60	0.00	0.00
Latvia	24.93	43.93	32.68	0.00	0.00
Liechtenstein	0.00	0.00	0.00	0.00	0.00
Lithuania	32.72	590.34	8.62	0.00	0.00
Luxembourg	2.00	206.00	62.00	0.00	0.00
Mexico	1010.23	904.92	35.95	0.00	0.00

	Weighted (SEN code=1)	Weighted (SEN code=2)	Weighted (SEN code=3)	Weighted (SEN code=4)	Weighted (SEN code=5)
Montenegro	0.00	0.00	0.00	0.00	0.00
Netherlands	178.15	470.19	0.00	0.00	0.00
New Zealand	191.30	824.44	748.86	0.00	28.87
Norway	90.38	1756.23	413.65	0.00	0.00
Panama	0.00	0.00	0.00	0.00	0.00
Peru	245.12	312.84	0.00	0.00	0.00
Poland	169.11	1061.16	0.00	0.00	0.00
Portugal	24.79	1322.16	196.60	0.00	0.00
Qatar	9.00	18.00	1.00	0.00	0.00
Republic of Korea	994.40	753.21	0.00	0.00	0.00
Romania	0.00	0.00	0.00	0.00	0.00
Russian Federation	2080.53	13009.70	156.79	0.00	0.00
Serbia	65.66	53.38	0.00	0.00	13.49
Singapore	17.44	216.90	182.35	0.00	0.00
Slovak Republic	170.89	558.30	18.97	767.53	0.00
Slovenia	40.47	31.63	65.73	0.00	0.00
Spain	1006.77	7140.80	4525.13	0.00	0.00
Spain (Andalusia)	101.35	1927.16	315.40	0.00	0.00
Spain (Aragon)	7.09	107.18	217.25	0.00	0.00
Spain (Asturias)	8.65	20.03	0.00	0.00	0.00
Spain (Balearic Islands)	17.61	26.93	83.53	0.00	0.00
Spain (Canary Islands)	9.38	74.97	97.89	0.00	0.00
Spain (Cantabria)	0.00	88.28	44.66	0.00	0.00
Spain (Castile and Leon)	11.03	268.55	175.58	0.00	0.00
Spain (Catalonia)	69.58	1580.45	1314.38	0.00	0.00
Spain (Ceuta and Melilla)	5.78	30.04	8.63	0.00	0.00
Spain (Galicia)	37.44	456.24	75.43	0.00	0.00
Spain (La Rioja)	8.28	38.20	33.23	0.00	0.00
Spain (Madrid)	83.90	1021.58	561.75	0.00	0.00
Spain (Murcia)	8.10	497.19	278.39	0.00	0.00
Spain (Navarra)	0.00	63.16	36.13	0.00	0.00
Spain (Basque Country)	25.86	199.33	168.09	0.00	0.00
Sweden	2627.73	0.00	731.91	0.00	0.00
Switzerland	64.30	343.61	532.19	0.00	0.00
Thailand	0.00	259.69	198.22	0.00	0.00
Tunisia	104.41	21.42	57.98	0.00	0.00
Turkey	337.96	494.63	664.78	0.00	0.00
United Arab Emirates (Dubai)	1.63	1.63	3.42	0.00	0.00
United Kingdom	2438.26	13482.24	1173.73	0.00	0.00
Scotland	239.78	1100.46	201.81	0.00	0.00
United States of America	15367.22	127485.82	21718.22	5970.96	0.00
Uruguay	14.48	34.00	18.10	0.00	0.00

Weighted (SEN code=1)	Weighted (SEN code=2)	Weighted (SEN code=3)	Weighted (SEN code=4)	Weighted (SEN code=5)
--------------------------------------	--------------------------------------	--------------------------------------	--------------------------------------	--------------------------------------

*Denmark code 4 within-school exclusion defined as students with dyslexia/-calculi

*Finland code 4 within-school exclusion defined as students with dyslexia

*Greece code 4 within-school exclusion defined as students with dyslexia

*Hungary code 4 within-school exclusion defined as students with partial skills deficiencies (dyslexia, dysgraphia, etc.)

*Iceland code 4 within-school exclusion defined as students with dyslexia

*Ireland code 4 within-school exclusion defined as students with dyslexia

*Slovak Republic code 4 within-school exclusion defined as students with dyslexia

*United States of America code 4 defined as students who are home-schooled