PISA for Development Strand C Technical Standards

Doc.: PISA-D (Strand C) TR Annex C Technical Standards_Draft.docx

February 2017

Produced by the OECD and the PISA-D contractors

Table of Contents

PISA FOR DEVELOPMENT STRAND C TECHNICAL STANDARDS

Purpose of document

The purpose of this document is to list the set of standards upon which the PISA for Development (PISA-D) Strand C data collection activities will be based, as is the case in the regular PISA project. These standards rely on the PISA 2018 standards but are adjusted for PISA-D Strand C given its differences from PISA. The main difference in Strand C is that it is planned to be implemented as a household survey with data collected through in-person interviews accompanied by a household paper-based questionnaire to be answered by the most knowledgeable person about the respondent. In following the procedures specified in these standards, the partners involved in the data collection activities contribute to creating an international dataset of a quality that aims at allowing for valid cross-national inferences.

The major difference between PISA/PISA-D Strand A and PISA-D Strand C is the target population. PISA and PISA-D Strand A rely on in-school 15 year-olds in grades 7 or above, who are thus referred to as students while the PISA-D Strand C target population includes 14-16 year-olds that are out-of-school broadly defined. The definition of out of school in PISA-D builds on the work of the UNICEF and UNESCO Institute for "Statistics All in School: Global Initiative on Out-of-School Children*".*[1] This effort identifies 5 dimensions of exclusion, of which PISA-D will cover those concerning 14-16 year old youth, notably, those that have not been to school, those who attended primary school and dropped out during primary school, those who completed primary school and may or may not have some secondary school but have already left the schooling system, and those in school attending primary education (that is in grades 6 or below). Thus, within this document, the PISA-D Strand C target population will be referred to as "Strand C target population" or "Strand C respondents". The population of Strand C is the complement of the population in Strand A to ensure that the entire cohort of 15 year olds is covered by PISA-D. At this point, however, it is worth noting that the goal of Strand C is not necessarily to provide a representative sample of the out of school population thus defined, but rather to develop an approach, tools and mechanisms that will allow this to be done in future rounds of PISA.

These standards for data collection and submission were developed with three major, and inter-related, goals in mind: consistency, precision and generalisability of the data. Furthermore, the standards serve to ensure a timely progression of the project in general.

- **Consistency**: Data should be collected in an equivalent fashion in all countries (see Mohler and Johnson (2010)), using equivalent materials that were translated and/or adapted as appropriate. Given consistent data collection (and sufficiently high response rates) and equivalent samples, results are likely to be comparable across countries. Thus, the test results in different countries would reflect differences in the performance of the youth measured, and will not be caused by factors which are un-related to performance.

---

[1] For more details see: http://www.uis.unesco.org/Education/Pages/out-of-school-children.aspx

- **Accuracy**: Data collection and submission practices should minimise spurious variation or error to the greatest extent possible. In the total survey error context, accuracy is the sum of the variance of an estimate and the various squared bias components. This holds for both systematic and random error sources, e.g. when the testing environment differs from one respondent to another, measurement errors, and processing errors such as when data entry procedures leave room for interpretation. An increase in accuracy relates directly to the quality of results one can expect. The more precise the data, the more powerful the (statistical) analyses, and the more trustworthy the results to be obtained, given there are no biases.

- **Generalisability**: Data are collected from specific individuals in uniform conditions and at a certain time interval. Individuals to be tested should be selected, and test materials and questionnaires etc. be developed in a way that will help to ensure that the conclusions reached from a given set of data do not simply reflect the setting in which the data were collected but hold for a variety of settings and are valid in the target population at large. Thus, collecting data from a representative sample of the population, for example, will be essential, but not sufficient, for the results to accurately reflect the level of literacy of the Strand C target population.

- **Timeliness**: Consistency, precision and generalisability of the data can be obtained in a variety of ways. However, the tight timelines and budgets in PISA-D, preclude the option of developing and monitoring local solutions to be harmonized at a later stage in the project. Therefore, the standards specify guidelines for data collection and data submission.

This document strives to establish a collective agreement of mutual accountability among countries, and of the international contractors towards the countries. This document details each standard in terms of quality assurance and a quality management plan designed to demonstrate that the standard has been met. While the terms quality assurance and quality control are sometimes used interchangeably, they relate to different aspects of quality. Quality assurance is most often associated with the processes and procedures that are put in place to make sure the survey is likely to meet its intended goals. Quality control, on the other hand, relates to the set of judgments that are made with regard to the suitability of the survey results in terms of their intended uses or applications.

Where standards have been fully met and data quality of the final database judged as appropriate, the international contractors will recommend to the OECD Secretariat that the data be included in the PISA-D Strand C database. Where standards have not been fully met or data quality questioned, an adjudication process will determine the extent to which the quality and international comparability of the data have been affected or whether additional analysis or evidence are necessary. The result of data adjudication will determine whether the data will be recommended for inclusion in the PISA-D Strand C dataset.

In principle, each dataset should be evaluated against all standards jointly. Also, it is possible that countries' proposed plans for implementation are not, for various and often unforeseen circumstances, actually implemented. Therefore, the final evaluation of a participant's adherence to these standards needs to be made with respect to the data as submitted since this is the definitive indication of what may appear in the released international dataset.

If any issues with attaining standards or data quality are identified, the international contractors initiate communication with the National Centre as soon as possible to give advice on resolving problems.

These standards are supposed to serve as benchmarks of best practices. As such, the standards are designed to assist National Centres and international contractors by explicitly indicating the expectations of data quality and study implementation endorsed by the OECD, the PISA Governing Board and the International Advisory Group (IAG), and by clarifying the timelines of the activities involved. The standards formulate levels of attainment, while timelines and feedback schedules of both the participating countries and the international contractors are defined in the PISA-D operations manuals.

The international contractors take responsibility for developing and implementing procedures for assuring data quality. Therefore, the international contractors mediate and monitor the countries' activities specified in this document.

Where the technical standards stipulate that deviations from the standards require agreement between participating countries and the international contractors, National Project Managers are asked to initiate the process of negotiation and to exercise all possible options in order to facilitate an agreement. Where agreement between National Project Managers and the international contractors cannot be reached, the OECD will adjudicate and resolve the issues. The OECD will also adjudicate any issues resulting from non-compliance with the technical standards that cannot be resolved between participating countries and the contractors.

There are three types of standards in this document, each with a specific purpose:

- **Data Standards** refer to aspects of study implementation that directly concern the data quality and its control. These standards have been endorsed by the Technical Advisory Group and wherever proportions or quantities are specified (for example, response rates), these have been reached through examination of research undertaken or have been reviewed by members of the Technical Advisory Group with the aim of minimising the effect of any potential bias in the data.

- **Management Standards** are in place to facilitate that all PISA-D operational objectives are met in a timely and coordinated manner.

- **National Involvement Standards** help to ensure that the internationally developed instruments are widely examined for cross-national, cross-cultural and cross-linguistic validity and that the interests and involvement of national stakeholders are considered throughout the study.

Format of the document

The standards are grouped into sections that relate to specific processes in the PISA-D Strand C. For every section, a rationale is given explaining why standards are necessary. The standards in each section consist of three distinct elements. First, there are the Standards themselves that are numbered and shown in shaded boxes. Second, there are Notes that provide additional information on the standards. The notes are listed after the standards in each section. Third, there are the quality control measures that will be used to assess if a standard has been met or not. These are listed at the end of each section. In addition, the standards contain words that have a defined meaning in the context of the standards. These words are shown in italics throughout the document and are clarified in the Definitions section at the end of the document, where the terms are listed alphabetically.

Scope

The standards are stated in the context of the future cycles of PISA in which countries will assess an out-of-school population using procedures established in the Strand C pilot. Once this pilot is complete, the PISA Governing Board will review and approve an updated set of standards for future out-of-school PISA assessments. Where appropriate, notes are added about the pilot project Main Survey, and any differences in the pilot project Field Trial are noted in Section 5.

Data standards

1.   Target population and sampling

Rationale: The sampling standards follow the objectives of Strand C, a pilot project with two main objectives for the sample:

- • To yield a large enough sample size to test the validity of the questionnaire and the assessment items and other studies planned for this Strand, including linking to PISA and PISA-D Strand A (reporting in the same scale).

- • To explore various approaches and evaluate various options to arrive at a recommendation for selecting and assessing a nationally representative sample of non-institutionalised 14 to 16-year-olds who are out of school as defined above in future cycles of PISA, for countries choosing this option.

For the pilot project, countries are allowed to use nonprobability sampling to satisfy the first objective but are required to design and select a probability sample to satisfy the second objective. The proportion in each strategy in national samples will be negotiated with the contractors, keeping in mind the objectives of Strand C as well as the affordability of the survey.

The set of standards developed for the pilot project will be modified at the conclusion of the Main Survey for use in future cycles of PISA, starting with 2021. Planning criterion for future PISA cycles based on accuracy (total survey error) and other constraints such as specific comparability criteria will be further developed through the pilot project.

The method of data collection will be in-person interviews, most likely in households or other agreed upon locations. Generally, costs associated with in-person interviewing are much higher than those in school samples, especially when the target sample is a rare or hard-to-sample population, as is the case for Strand C. Therefore, a critical component of this pilot project is to arrive at sampling plans that minimize costs to countries, given the objectives of Strand C.

Each participating country will appoint a National Sampling Manager (NSM) to take responsibility for all sampling-related activities of PISA-D Strand C, including sample design, selection, sample monitoring, and providing the standard set of sampling-related data to the International Contractor. This is a new position that does not exist in PISA but must exist in PISA-D Strand C. The NSM should have experience in household surveys, albeit not necessarily focused on education. The NSM will help define as clearly as possible the target population, which is necessary to ensure that adequate steps are taken to accurately cover the population(s) of interest in the sampling process in each participating country.

The sampling frame is the source from which the sample is selected at the given stage of sampling, and so the quality of the sampling frame (coverage rate and access to auxiliary information) affects the quality of the sample. Therefore, sampling frames must meet a minimum set of quality standards to ensure that adequate and accurate information is available for carrying out sampling, data collection, weighting, and nonresponse bias analyses. It is also important that exclusions be clearly specified and limited as much as possible so that no extensive biases are introduced as a result of under-coverage of the population. Exclusions need to be documented thoroughly and transparently to assess the representativeness of the sample in PISA-D Strand C.

The psychometric analysis for the Main Survey of Strand C requires adequate sample sizes for each tested language in a participating country. The achievement of adequate minimum sample sizes for each tested language will enable the Consortium to estimate separate population models. Population modelling is a critical step in obtaining appropriate proficiency values that will be used in describing the distributions of skills in a country and in reporting national and subpopulation data.

In future PISA cycles, the goal is that each individual in the target population will have a non-zero probability of selection resulting from the application of established and professionally recognised principles of scientific sampling. In the pilot project, a large portion of the sample should be selected from a probability-based sample (at least 50 percent). Probability-based scientific sampling is essential for two main reasons. First, probability sampling encompasses a set of designs that leads to a variety of unbiased sampling approaches that allow analysts to generalise the results to the target population. Second, measures of precision related to survey estimates (i.e. standard errors, margins of error, confidence intervals) can be computed only under a probability design. Hence, statistical tests for differences between survey estimates are possible only for the portion of the pilot study sample selected under a probability-based design.

Within the context of total survey error and maximising the accuracy of the survey results, the aim in developing the probability sample design and the selection process is to achieve the maximum precision possible for a given sample size, while limiting the costs of data collection to a fixed budget. The Strand C probability-based sample design will be a stratified multi-stage clustered area sample. Deviations from the core design are expected because some countries are geographically small, and therefore will have less clustering and fewer stages of sampling. Also, some countries will have lists of dwelling units (including group quarters) or individuals already available from national registries. The general approach will allow for flexibility in the sample design and will be adaptable to each country's best sampling scenario that strives to minimise biases, such as from non-coverage and nonresponse, and produces high precision estimates, to improve comparability for a fixed cost minimised costs to the countries.

For the probability sample, the units selected at each stage of sampling must be verifiable, i.e. the same sample would be selected if the same sampling procedures were repeated with the same *random seed* used to trigger the selection. Errors in sample selection will produce a bias in survey estimates; therefore, the international contractor must approve each country's sample design plan and countries must conduct checks at each stage of sample selection.

Under ideal situations in a probability sample, every eligible member of the target population (as defined for Strand C) would have a non-zero probability of selection in a national sample, would be located and would agree to participate in the study. In practice, these circumstances are not realised in any survey population, which results in potential for bias due to nonresponse and non-coverage.

While there have been studies (such as by Keeter et al. (2000) and Curtin, Presser and Singer (2000)) suggesting that nonresponse rates are not as strongly related to nonresponse bias as previously thought, it is well understood that when response rates are low, there is a greater risk for nonresponse bias (Groves and Peytcheva, 2008). The extent of nonresponse bias depends on many survey conditions, including the differential impact that the likelihood of response has on the bias of each of the survey outcomes. It is, therefore, critical to take steps to reduce nonresponse bias during data collection, and to evaluate the potential for nonresponse bias, as a quality check on the estimates, at the conclusion of the data collection. Similarly, under-coverage bias (due to exclusions) can be substantial if the under-coverage rate is high (a large part of the target population is excluded from the frame) and the difference in proficiency levels between the eligible 14 to 16 year olds included in the sample and those excluded from the frame is relatively large. The maximum allowable exclusion rate is set at 5% to guard against high under-coverage bias in estimates. Given the relationships between bias and under-coverage and response rates, countries must keep the exclusion rates low and implement procedures to reduce the potential for nonresponse bias and attain high response rates.

There are several ways to reduce the potential for nonresponse bias. First and foremost is to plan and implement field procedures that achieve a high level of cooperation. It is also critical to monitor the distribution of the sample during data collection to ensure steps are taken to reduce the potential for bias as much as possible. As discussed in Groves (2006), as nonresponse rates increase, one needs to actively seek auxiliary data to reduce the impact of nonresponse propensities on the survey estimates. These auxiliary variables can then be used in weighting adjustments for the purpose of reducing nonresponse bias. See Stoop, et al. (2010) and Bethlehem, et al. (2011) for more information.

Respondent incentives have been shown to be effective for improving response rates without affecting the respondent's performance. As a result, the use of incentives can potentially reduce bias in the estimates. See, for example, Mohadjer et al. (1997) and Singer (2002). Incentives have a tendency to attract certain groups of initial non-respondents (see also Couper, et al. (2008)). Because the effectiveness of incentives will vary by country, each country should choose what works best for its situation.

Close monitoring of the data collection is critical to the success of the data collection effort. The monitoring process can help countries identify potential shortfalls in the sample, problems in achieving the desired response rate and potential for nonresponse bias in the collected sample. Continuous monitoring will allow countries to employ procedures to address these problems during data collection, such as possibilities for intervention when necessary (measurement errors, nonresponse, fabrication patterns and paradata to reflect these), while it is still possible to meet goals associated with sampling and data quality.

Standard 1.1 The National Centres must assign a **National Sampling Manager (NSM)** to oversee sampling-related activities and communicate with the International Contractor sampling group**. The NSM** must be distinct from the NPM and have the following qualifications:

- Experience in scientific sample design and selection of household probability samples;

- Familiarity with sampling hard-to-reach populations or at least methods developed for that purpose;

- Familiarity with weighting and nonresponse bias analysis procedures; and

- Proficiency in English.

Standard 1.2 The *PISA-D Strand C target population* consists of 14 to 16 year-olds in-school at grades 6 and lower, and non-institutionalised 14 to 16 year olds who are out-of-school.

Standard 1.3 The sampling frame should include 95% or more of the core *PISA-D Strand C target population*. That is, the under-coverage rate, combined over all stages of sampling, should not exceed 5%.

Standard 1.4 The minimum sample size requirement is 2 700 completed cases for the Main Survey. For the pilot project Main Survey, ideally, a large portion of the sampled cases will come from the probability component of the Strand C sample. Countries should achieve a minimum completed sample size of 2 700 respondents for their main language, with additional completed cases needed as proportionate to the population. This sample characteristics does not allow separate reports produced by language. Countries will have an option for oversampling subgroups of interest. This will be in addition to the required sample size.

Standard 1.5 A completed case is one that contains responses to key background questions and a sufficient number of assessment items.

Standard 1.6 In future cycles, the goal is that each individual in the target population will have a known non-zero probability of selection resulting from the application of established and professionally recognized methods for scientific sampling. In the pilot project Main Survey, a large portion of the sample should be selected from a probability-based sample. The allocation of the Main Survey sample to the probability/nonprobability-based components will be variable across countries. The International Contractor will work closely with the participating countries to arrive at a sample design that maximizes the size of the probability-based component without jeopardizing the success of this pilot study in various countries.

Standard 1.7 For future cycles, the core design should be a self-weighting design of households (or of individuals, for countries with person registries) within approved sampling domains used to identify areas with high and low concentration of the target population. A self-weighting design is typically achieved when each sample household (or individual, if not sampling dwelling units) has an equal probability of selection.

Standard 1.8 For countries with a screener questionnaire, the goal is to select all eligible individuals in the target population within the household.

Standard 1.9 The sampling frame(s) at each stage of selection must be up to date and must include only one record for each sampling unit of the target population. The sampling unit at each stage of selection must be clearly defined. Special care must be taken to eliminate duplicate records in the situation where lists have been combined to create a sampling frame. Countries should assess the extent of duplication and the proportion of out-of-scope units on the frame and if necessary, develop a plan to correct these problems.

Standard 1.10 Countries must evaluate and develop a plan to address any under-coverage in the frame that was not addressed in the documentation of country-specific exclusions. To address under-coverage in the local area sampling frame, missed structure and hidden dwelling unit procedures can be implemented. The International Contractor will provide countries with a description of these procedures.

Standard 1.11 No substitution of sample units at any sampling stage (e.g., youth, dwelling units, enumeration areas) will be allowed in the probability sample.

Standard 1.12   For sample designs involving a screener questionnaire, eligible individuals must be identified from within households using a fully enumerated grid of household members.

Standard 1.13   A minimum overall response rate of 70% is the goal. The overall response rate will be computed as the product of the response rates for the stages included in data collection. All response rates must be weighted by the household base weight, in the case of a screener response rate, and by the person base weight, in the case of a BQ or assessment response rate.

Standard 1.14   Each participating country will provide completed quality control monitoring forms during each month of data collection to the International Contractor. The report must contain the number of cases completed, the number of cases worked, response rates and expected yield.

Standard 1.15   At the end of data collection, each country will submit its results to the Sample Design International File (SDIF). The contents of the SDIF will include sample selection data for each sampled unit, including sampling strata, probabilities of selection, and auxiliary variables for weighting adjustments.

Note 1.1   Age is defined as age at the time of the interview. For surveys with a screener questionnaire, age is determined when the screener is conducted. For countries with registries, age at the midpoint of the data collection period will be used to define the sampling frame of age eligible individuals. For example, suppose 1 December is the midpoint of data collection; then the eligibility of the 14 to 16 year old individuals will be based on their status as of 1 December.

Note 1.2   All youth are to be included regardless of citizenship, nationality or language.

Note: 1.3   Exclusions should be made only because of operational or resource considerations, for instance, excluding individuals in hard-to-reach areas. The country should try to identify exclusions before sample selection, to the extent possible.

Note 1.4   The sampling frame at each stage should contain any information necessary for sample design, sample selection and estimation purposes, and locating individuals

Note 1.5   A multi-stage cluster sample is necessary when a complete listing of eligible individuals or *housing units* is not available. In multi-stage cluster sampling, geographic clusters are formed and selected, giving each cluster a probability of selection. Within sampled clusters, a local registry of individuals may be obtained. If no person registry is available, all dwelling units (including group quarters, sometimes referred to as collective dwelling units) are listed and a sample of dwelling units is subsequently selected, giving each dwelling unit probability of selection. Within selected dwelling unit, the eligible individuals are identified and selected.

Note 1.6   The International Contractor will provide countries with information about dwelling unit listing procedures used to create local area frames.

Note 1.7   The initial sample size should be adjusted to account for expected nonresponse and ineligibility. For countries with a screener, sample size goals should be constructed for the screener, to account for ineligibility and screener nonresponse, as well as nonresponse to the background questionnaire and the assessment.

Note 1.8   For the pilot project Main Survey, the sample design at each stage of sampling for the probability sample component, will allow an overall probability of selection to be derived for each individual. Non-probability designs, such as link-tracing *network sampling* (snowballing, respondent driven sampling), *quota sampling* and the *random route* approach, are only allowed in the pilot project Field Trial and Main Survey. Countries have the option to increase the sample size to achieve psychometric goals using non-probability designs to obtain enough cases for item parameter estimation. See Tourangeau et al (2014) for a description of various methods to handle hard-to reach populations.

Note 1.9   For units such as dwellings or *Basic Dwellings*, household and *country of usual residence* must be defined clearly by the countries to determine whether or not the sampling unit belongs to the target population.

Note 1.10   A full enumeration grid of household members must be completed to allow for a thorough and systematic approach to determining the eligibility of household members. If full enumeration is not used, there is a greater chance of missing eligible household members. Full enumeration also facilitates the collection of specified variables (e.g. gender) for each individual, which can be used in a weighting adjustment for nonresponse.

Note 1.11   At the end of each stage of selection, countries will complete a sample selection quality control form to help the International Contractor verify that the sample selection process was conducted accurately.

Note 1.12   At the end of sample selection, each country will construct a survey control file for their own sample management purposes. The survey control file will have a record for each sampled household for countries with household samples, and will have a record for each sampled individual if individuals are selected directly from a registry. At a minimum, the survey control file must include a unique ID for each record, country ID, location information, identification of sample analysis groups (e.g., language areas), population strata and sampling units at different stages, probabilities of selection (for the probability sample), and other sampling or frame information. Not all survey control file contents will need to be loaded into the *study management system* to support the data collection effort.

Note 1.13   Data from all countries with a minimum response rate of 70% will generally be included in international indicators and reports unless sample monitoring activities and/or nonresponse bias analyses conducted by the International Contractor indicate serious levels of bias in the country data. Results from countries with response rates below 50% will not be published unless OECD Secretariat is provided with evidence that the potential bias introduced by the low response rates is unlikely to be greater than the bias associated with response rates of between 50% and 70%.

Note 1.14   Sample monitoring must begin the first week of the data collection process so that potential problems can be identified and corrected as soon as possible.

Note 1.15   Each participating country may opt to provide a "modest" incentive to obtain respondent co-operation, such as a monetary or non-monetary incentive (e.g. pen, notepad, candy, mug, voucher, gift certificate. However, the planned incentive must be approved by the International Contractor with costs covered by the participating countries. Countries that are interested in using incentives in the MS should experiment with various options in the FT.

Note 1.16   A layout for the Sample Design International File (SDIF) will be provided by the international contractors. Data for the SDIF will be provided by the country through the Data Management Expert (DME).

Quality Control

The quality control process for sampling-related activities is mainly facilitated by the countries completing forms or data files for the purpose of helping to ensure that the standards are followed and the total survey error is reduced.

Countries should provide the international contractor with the name and contact information of their NSM.

Sample selection quality control forms. The sample selection QC forms provide summary information from the countries to allow the International Contractor to verify each stage of selection using Excel templates.

Sample monitoring quality control forms. Countries will submit sample monitoring quality control forms to the International Contractor to help countries identify shortfalls in the sample, problems in achieving the desired response rate, and the potential for nonresponse bias in the collected sample. Continuous monitoring of the sample will allow countries to employ procedures to address these problems during data collection, while it is still possible to meet goals associated with sampling and data quality.

Sample design international file (SDIF). Countries are responsible for providing case information according to the standard layout to form the basis for post-data collection statistical processing and verification of the sample.

Before and during data collection, countries are required to evaluate and take steps to reduce the potential for nonresponse bias. After the Main Survey data collection, countries will conduct a basic nonresponse bias analysis (NRBA) which includes, at minimum, selecting the variables for weighting, and comparing the weighted response rates by subgroup. The international contractor will perform an extended NRBA. Countries will provide the data necessary for conducting the extended NRBA to the international contractor.

2.   Adaptation of tests, questionnaires and manuals

Rationale: To validly compare performance across countries, all assessment instruments and other survey materials have to be as equivalent as possible. Therefore, assessment instruments and questionnaires as well as all sets of instructions and the procedures of data collection have to be equivalent. To achieve this goal, other individuals who play a key role in the data collection process, i.e. interviewers, should receive the same information and training in all participating countries.

Standard 2.1 The majority of test items used for linking with PISA or PISA-D Strand A are administered unchanged from their previous administration. The computer-based versions include instructions as to the appropriate response mode for each item and may require some minor revision based on the functionality of the tablet delivery system.

Standard 2.2 All assessment instruments should be psychometrically equivalent to the source versions. Agreed upon adaptations to the local context are made if needed, following the guidelines and procedures provided by the international contractors.

Standard 2.3 The questionnaires should be equivalent to the source versions. Agreed upon adaptations to the local context are made if needed, following the guidelines and procedures provided by the international contractors.

Standard 2.4 Manuals and scripts are equivalent to the source versions. Agreed upon adaptations to the local context are made if needed.

Quality Control

Agreed upon adaptation to manuals and scripts using method(s) specified by the international contractor

A Questionnaire Adaptation Spreadsheet (QAS), in which adaptations to questionnaires are documented

Test Adaptation Spreadsheets (TAS, for paper and computer instruments), in which adaptations to assessment units, orientation and help files and common booklet parts are documented

Item and scale statistics generated by the international contractors (assessment materials and questionnaires)

3. Translation of assessment instruments, questionnaires and manuals

Rationale: To be able to compare the performance of respondents across and within countries, the linguistic equivalence of all materials is central. While Standards 2.1 to 2.4 serve to ensure that equivalent information is given to respondents in all countries involved, in general, the following Standards 3.1 and 3.2 emphasise the importance of language. Again, the goal is to ensure that literacy will be assessed, and not variations of information caused by differences in the translation of materials.

Standard 3.1 The following documents are translated into the assessment language in order to be linguistically equivalent to the international source versions.

- All assessment instruments

- All questionnaires

- The interviewer Procedures Manual (and training scripts)

Standard 3.2 Unless otherwise agreed upon, manuals are translated/adapted into the assessment language to make them functionally equivalent to the international source versions.

Note 3.1 This standard apply to instruments that are in a language that is administered to more than 10% of the target population.

Note 3.2 It is not required to translate the Quality Control and Management Manual if the National Centre staff who will be using it can read and understand English.

Quality Control

Agreed upon Translation Plan developed in accordance with the specifications in the PISA-D operations manuals where the Translation Plan would require double translation by independent translators.

Agreed Upon Questionnaire Adaptation Spreadsheet (QAS)

Test Adaptation Spreadsheets (TAS), in which adaptations to assessment units, orientation and help files, common booklet parts and coding guides are documented. Adaptations will be checked for compliance with the PISA-D Translation and Adaptation Guidelines by international verifiers, and the verifiers' recommendations will be vetted by the translation referee.

Verifier Reports (statistics generated by the TAS, in combination with a short qualitative report)

Submitted final materials as used in the study

Item and scale statistics generated by the international contractors (assessment materials and questionnaires)

4. Testing and support of national software versions

Rationale: Countries participating in PISA-D Strand C will be primarily responsible for resolving system-related operational issues in their countries, including hardware issues and provision of technical support to interviewers. Prior to data collection, countries must thoroughly test and validate the national software releases that are used to deliver the PISA-D Strand C computer-based instruments (Strand C instruments are delivered via tablet computers).

Standard 4.1   The National Centres must assign a **National Information Technology Coordinator (ITC)** to oversee activities related to the national delivery systems and communicate with the International Contractor information technology group. The **ITC** must have the following qualifications:

- familiarity with the PISA-D Strand C computer system requirements applications and training materials;

- familiarity with all national software standards and procedures; and

- attendance at all relevant training sessions to become familiar with the computer-based assessments and understand the way in which the interviewers are required to interact with the interface.

Standard 4.2   The international contractors must test all national software versions prior to their release to ensure that they were assembled correctly and have no technical problems.

Standard 4.3   Once released, countries must test the national software versions following testing plans from the international contractors to ensure the correct implementation of national adaptations and extensions, display of national languages, and proper functioning on tablets.

Standard 4.4   Each country should have a designated PISA-D Strand C helpdesk with contact information provided to each of its interviewers. The country helpdesk staff must:

- be familiar with the PISA-D computer system requirements applications and training materials,

- be familiar with all national software standards and procedures; and attend the *interviewer* training sessions to become familiar with the computer-based assessments.

Note 4.1   Errors found during testing should be promptly communicated to the international contractors using agreed-upon problem reporting procedures. These procedures require that testing results are shared with the international contractors in order to monitor the quality of the instruments.

Quality Control

Detailed testing plans

Review of testing results

5.   Field Trial participation

<u>Rationale:</u> The Field Trial gives countries the opportunity to try out the logistics of their data collection procedures and allows the International Contractors to make detailed analyses of the items so that only suitable ones are included in the Main Survey. The minimum sample size required for the Field Trial in Strand C is based primarily on psychometric considerations. Without an adequate number of participants responding to each of the field-tested items in each assessed language, the International Contractor would be unable to evaluate the fit of the item parameters across the languages used in each country.

> **Standard 5.1**   The minimum sample size requirement for the Field Trial is 1 200 completed cases for the major language.
>
> **Standard 5.2**   For the Field Trial, countries will design and select a probability-based sample (not required to be a representative sample) to test out the logistics, as well as implementing a non-probability design to obtain enough cases for item parameter estimation. See more details related to probability and non-probability sampling in Section 1.

6.   Security of the material

<u>Rationale:</u> The goal of the PISA-D Strand C assessment is to measure the literacy levels in the content domains of reading and mathematics. Prior familiarisation with the test materials will heavily degrade the consistency and validity of the data. In the extreme case, the results would only reflect how well participants are able to memorise the test items. In order to be able to assess the competencies obtained during schooling rather than short-term learning success, and to make valid international comparisons, confidentiality is extremely important.

> **Standard 6.1**   PISA-D Strand C materials designated as secure are kept confidential at all times. Secure materials include all assessment materials, data, and draft materials. In particular:
>
> - no one other than approved project staff and respondents (during the interview) is able to access and view the assessment,
>
> - no one other than approved project staff will have access to secure PISA-D data and embargoed material, and
>
> - formal confidentiality arrangements will be in place for all approved project staff.
>
> **Standard 6.2**   PISA-D interviewers must be trained on the importance of data security using guidelines provided by the contractors.
>
> **Standard 6.3.**   Interviewers must maintain the security of their tablet computers, keeping passwords separate from the tablet computer and using the tablet computer only for PISA-D-authorised activities.

Note 6.1   Interviewers must read and sign an agreement to keep tablet computers and survey materials secure in respondent homes, cars, and other locations, to prevent theft.

Note 6.2   Interviewers must minimise the number of paper records in their possession, by returning materials with respondent identifiers to the survey institute upon completion of each case.

Quality Control

Countries will be required to provide the details of their proposed data collection security plan.

Confidentiality form signed by interviewers agreeing to the security measures stated in Standard 6.3.

Survey Operations Data Collection Form. Starting two months before the data collection period until two months after, countries will be required to submit this form monthly to inform the discussion during monthly QC calls with the International Contractor.

FT and MS Final Survey Operations Report should thoroughly document the actual survey experiences in each country. This report, produced by international contractors, is based on information collected during the survey operations quality control process. The report should identify the aspects of the survey that worked well and also include an explanation of problems encountered. International contractors will provide an outline prior to the completion of the report.

7.   Assembling and printing paper-based questionnaires

Rationale: Variations in assembly and print quality may affect data quality.

Standard 7.1   The cover page of the PISA-D paper-based household questionnaire must contain all information as specified by the PISA Governing Board.

Standard 7.2   The layout and formatting of the paper-based household questionnaire is equivalent to the source versions, with the exception of approved changes made necessary by national adaptations.

Note 7.1   The cover page of all PISA-D instruments should contain all information necessary to identify the material as being part of the data collection process for PISA-D, and for checking whether the data collection follows the assessment design, i.e. whether the mapping of the respondent on the one hand, and the household questionnaire, on the other, have been correctly established.

Quality Control

Submitted sample of print-ready questionnaire for optical check

8.   Data Collection

Rationale: Certain variations in the testing procedure are particularly likely to affect test performance. Among them are timing of the assessment, the administration of test materials and support material like rulers and calculators, the instructions given prior to test administration, and the rules for excluding respondents, etc. A list of relevant test conditions is detailed in the Interviewer Procedures Manual. To help ensure that the data are collected consistently and in a comparable fashion for all respondents, it is very important to standardise the data collection process to the extent possible.

The goal of the survey is (i) to yield a large enough sample size to test the validity of the background questionnaire and the assessment items and (ii) to arrive at a recommendation for identifying and assessing a nationally representative sample of 14-16 year olds that are out of school or in grades 6 and lower in future cycles of PISA-D.

The interviewer plays a critical role in data collection. Special consideration is therefore given to the training of the interviewers, ensuring that as little variation in the data as possible is caused by random or systematic variation in the activities of interviewers. Data collection shall begin immediately following interviewer training to maximize interviewers' using practices and procedures detailed during training

and to effectively channel the momentum built up during training. All study materials required to conduct the interviews as well as materials needed for administrative tasks shall be made available to the interviewers before they receive their work assignment.

An important part of the assessment situation relates to the relationship between interviewers and respondents. Strict objectivity of the interviewer is fundamental in collecting data that reflect the level of literacy obtained, and that are not influenced by factors un-related to literacy such as third-party presence. The results based on these data will be representative for the target population surveyed.

Standard 8.1   All interviews shall follow international procedures as specified in the *Interviewer Procedures Manual*, particularly the procedures that are related to:

- test administration and timing,

- test administration conditions, and

- respondent tracking.

Standard 8.2   Introductory material explaining the purpose of the PISA-D survey, shall be provided to each selected respondent or household prior to the interview.

Standard 8.3   The administration of the PISA-D interview components shall be conducted in person and in the following sequence - the screener, the background questionnaire and lastly, the assessment. All components should ideally be completed in the same visit.

Standard 8.4   In cases where the respondent does not speak the language(s) of administration, the use of interpreters is required for the administration of the background questionnaire as a way to limit the occurrence of literacy related non-response to the background questionnaire to the extent possible, if no bilingual interviewer is available to administer the interview. A member of the household can act as an interpreter during the administration of the background questionnaire and the parent questionnaire, but not the Exercise.

Standard 8.5   Make sure the selected individual is interviewed and assessed. No proxies are allowed, not even for the BQ.

Standard 8.6   Instruments (including the screener, background questionnaire and the assessment) shall be administered in all official languages as determined by each country.

Standard 8.7   Countries will employ a sufficient number of interviewers so that the maximum assignment size for an interviewer is 40 completed assessments per month.

For the Probability Sampling Component:

Standard 8.8   Initial contact with sampled households or individuals shall be in person with at least four contact attempts made before the case is coded as a non-contact. Additional efforts may be needed to keep the rate of noncontacts as low as possible.

Standard 8.9   The identification of target population members within a sampled household requires the administration of a set of questions (hereafter called the screener) that should be done by the interviewer. All eligible individuals within a sampled household should be selected for the survey.

Standard 8.10   The interviewer shall document details of each contact attempt with the household or respondent in a Record of Contact. Interviewers shall enter Record of Contact information each time a contact attempt is made, including the attempt that results in a completed interview.

Standard 8.11   The interviewer shall document details of each case that receives a non-interview disposition code because of refusal or other nonresponse.

Standard 8.12   The background questionnaire and the assessment must be administered to the selected individual; no substitution is permitted. However, the use of an interpreter is acceptable. The interpreter can be, for example, someone from the community, a professional interpreter, or someone in the respondent's household. The interpreter's only involvement shall be in translating the question to the respondent and translating the response to the interviewer. The interpreter shall not act as a proxy respondent.

Standard 8.13   Each country shall develop a "best practices" strategy for maximising cooperation and that strategy should be documented and used in training of staff. All countries shall implement effective refusal conversion strategies.

For the Non-Probability Sampling Component

Standard 8.14   For the pilot project, the non-probability approach of link-tracing may be used.

Standard 8.15   A limit shall be set on the number of peers that each participant can recruit to strive toward a broad representation of the target population. For this pilot study, each household and participant can recruit up to two potential eligible youths in their neighborhood. This limit can be relaxed as needed.

Standard 8.16   Population membership of new recruits for study participation shall be verified.

Note 8.1   Specific procedures for the verification of population membership of new recruits for study participation will be developed in conjunction with participating countries.

Note 8.2   Details of each contact attempt are documented in the record of contacts on the case folder.

Note 8.3   Details of each case that receives a non-interview disposition code are documented in the Non-Interview Report Form.

Note 8.4 Link-tracing can be implemented through selected households. The household representative or the eligible youth can refer to or recruit subjects within their neighbourhood (ideally within boundaries of the selected enumeration area) who are 14 to 16 years olds and either out-of-school or in-school below grade 7. To supplement, countries may also consider getting references or recruits through schools in the selected enumeration areas. For example, for out-of-school youth, schools with grade 7 can be approached to provide lists. For in-school youth, administrative lists can be requested of schools with 5th or 6th grade. The subjects who are referred to or recruited do not need to be selected randomly. However, they shall represent as much as possible 14-16 year-olds that are out-of-school and 14-16 year-olds that are in school but below grade 7.

Note 8.5 Countries could offer incentives (monetary or non-monetary) to respondents for participating in the study. The costs for these incentives are covered by the participating country. Countries interested in offering incentives in the Main Survey should pilot various offerings in the Field Trial, if possible.

Quality Control

Survey Operations Data Collection Form. Starting two months before the data collection period until two months after, countries will be required to submit this form monthly to inform the discussion during monthly QC calls with the International Contractor.

Monthly QC calls with National Centre staff. Starting two months before the data collection period until two months after, countries will be required to conduct a QC call with the International Contractor to discuss information provided on the Survey Operations Data Collection Form.

9.   Materials Development and Training Support

Rationale: Meeting the standards specified in this section will ensure that all data collection staff is thoroughly trained in the administration of the PISA-D Strand-C interview and assessment. To validly compare performance across countries and over time, it is of utmost importance to train data collection staff to achieve the highest quality of data using standardised procedures.

Standard 9.1   Countries shall develop interviewer materials that are required for the successful completion of the data collection effort from international versions provided by the international contractor.

Standard 9.2   International Contractors will assist each country to develop training materials on country-specific topics such as the procedures for locating households and respondents, screening questions and administrative procedures.

Standard 9.3   Countries shall develop a Supervisor Manual that field supervisors shall use as a reference document and guide to oversee interviewers and production implement administrative procedures and quality control measures.

Standard 9.4   All interviewer training sessions shall be scripted to ensure consistency of presentations across training sessions and across countries. Failure to do so could cause errors in data collection and may invalidate the results.

Standard 9.5   Qualified contractor staff shall conduct trainer training sessions with NPMs or their designees on PISA-D data collection materials and procedures. Countries shall conduct in-person interviewer trainings which must include sufficient project-specific training and cover General Interviewing Techniques (GIT).

> **Standard 9.6** NPMs or their designees shall participate in train-the-trainer sessions conducted by qualified contractor staff. This facilitates standardisation of training delivery to interviewers, allows trainers to become familiar with PISA-D materials and procedures, and informs trainers of their responsibilities for conducting the PISA-D training.
>
> **Standard 9.7** Qualified interviewers shall attend an in-person training which includes General Interviewing Techniques (GIT) as well as the project specific training on the concepts, instruments and procedures related to PISA-D. The in-person training is designed to maximise trainees' involvement and participation in the training and to provide opportunity for supervisory staff to observe and evaluate trainee performance.
>
> **Standard 9.8** Countries must evaluate the performance of interviewers during training, including their interviewing and gaining cooperation skills, use of the tablet, and comprehension of study protocols and procedures.
>
> **Standard 9.9** If there are multiple languages for the administration of the questionnaire and assessment within a country, interviewers must be trained in each language in which they will be conducting the interviews and assessment.

Note 9.1   The training package will include an Interviewer Training Guide with scripted training sessions, handouts, exercises and answer keys, scripted role plays, and a home study guide for interviewers.

Note 9.2   Field Supervisor refers to an individual appointed by the National Centre who is responsible for overseeing the work of data collectors in a particular region (or a subdivision of an area).

Note 9.3   The Interviewer Training Guide provides an overview of survey research with examples of survey questions and conventions for recording answers. The training includes important concepts related to gaining respondent cooperation and refusal aversion techniques, informed consent, data confidentiality, and standards and ethics in survey research. Interviewers will be trained on administering the survey and receive country –specific training on administrative procedures instituted by the National Centres.

Note 9.4   An international version of a Supervisor Manual will not be available due to the variability expected in supervising field work in each country. International contractors will be available to assist countries in developing the Supervisor Manual as needed.

Note 9.5   Countries must administer the Observation Form to each interviewer at the end of training.

Quality Control

Agreed upon Training Plan developed in accordance with the Interviewer Procedures Manual which details the operationalization of the Survey Operations standards.

Participation in trainer training sessions conducted by qualified contractor staff.

The international contractors will develop an evaluation form to be used to evaluate and document trainees' performance.

Interviewer training report submitted at the end of each training session conducted for interviewers.

10.   Field Management

Rationale: The skills of the data collection staff significantly affects the quality of the data collected. Therefore, ample time shall be devoted to thoughtful planning of the field staff resources required. Consideration shall be given to a variety of issues, including the number of field staff and their desired characteristics. Close supervision and mentoring of the data collection staff is required to produce the best quality data.

Each participating country will appoint a National Survey Operations Manager (NSOM) to take responsibility for all survey operations activities of PISA-D Strand C, including training materials preparation, interviewer recruitment and training, and field operations management.

> **Standard 10.1**   Diverse factors must be considered when determining the number of field staff required for a successful data collection effort. These factors include:
>
> - type of sampling frame(s)
> - sampled geographic area(s)
> - travel time to assignments,
> - typical hours per completed case for similar national surveys
> - interview length
> - length of the field period
>
> **Standard 10.2**   Countries are responsible for hiring the required number of interviewing staff with the desired characteristics.
>
> **Standard 10.3**   Countries shall monitor interviewer attrition and have a plan for hiring additional interviewers if needed.
>
> **Standard 10.4**   Various applicant qualifications shall be considered during the hiring process to ensure a qualified and competent data collection staff. These include interviewing experience, availability to work when respondents are most likely to be home, access to transportation and skills (*e.g.* language, communication)
>
> **Standard 10.5**   Adequate time shall be allotted for data collection staff recruiting and hiring. Interviewer recruiting and hiring shall commence with enough lead time prior to data collection to conduct all steps of the process, including identifying, interviewing and hiring candidates.
>
> **Standard 10.6**   The successful implementation of data collection requires survey management staff with clearly defined roles and responsibilities.
>
> **Standard 10.7**   Data collection shall be very closely monitored at all stages using observations, automated reports, validation, regular telephone calls with field staff, and review of data from completed interviews.

Quality Control

Survey Operations Planning Report. Countries are required to submit this report six months before the start of the data collection period to provide information to the International Contractor about the planning for survey operations activities.

Survey Operations Data Collection Form. Starting two months before the data collection period until two months after, countries will be required to submit this form monthly to inform the discussion during monthly QC calls with the International Contractor.

Monthly QC calls with National Centre staff. Starting two months before the data collection period until two months after, countries will be required to conduct a QC call with the International Contractor to discuss information provided on the Survey Operations Data Collection Form.

National Centre Quality Monitoring, in which the National Centre establishes and monitors the procedures stated above to provide technical support to the interviewers.

11.   Quality monitoring

Rationale: To obtain valid results from the assessment, the data collected have to be of high quality, i.e. they have to be collected in a consistent, reliable and valid fashion. The goal of fieldwork quality control monitoring is to minimize interviewer variance and bias which is implemented by the field management team (supervisors and field managers). The field management team is responsible for quality control activities at the regional and national level.

Standard 11.1   Each country shall develop and implement procedures to monitor the in-field performance of the interviewers. Monitoring can be done through methods such as in-person observation or audio recording of the interview.

Standard 11.2   Each country shall develop and implement procedures to verify 10% random subset of each interviewer's finalised work, including cases finalised as nonresponse. Validation procedures may include telephone calls or visits to sampled households by field management staff. Field interviewers can assist with in-person validation cases finalized by other interviewers as needed.

Standard 11.3   If an interviewer's work is found to be suspect, 100% validation of his/her other cases shall be conducted. Any falsified work shall be re-fielded and completed by another interviewer.

Standard 11.4   Interviewers shall review their hard-copy materials such as the case folder and parent questionnaire before finalising a case. Similarly, interviewers shall ensure that the instruments have been completed or the case is finalised before returning hard-copy materials to the survey institute.

Standard 11.5   The survey institute shall review all completed hard-copy materials and automated interview data to ensure that they meet the project standards for data.

Note 11.1   A failure to meet the Quality Monitoring standard in the Main Survey will lead to a significant lack of Quality Control data for other standards.

Quality Control

Survey Operations Data Collection Form. Starting two months before the data collection period until two months after, countries will be required to submit this form monthly to inform the discussion during monthly QC calls with the International Contractor.

Monthly QC calls with National Centre staff. Starting two months before the data collection period until two months after, countries will be required to conduct a QC call with the International Contractor to discuss information provided on the Survey Operations Data Collection Form.

12.   Data submission

Rationale: The timely progression of the project, within the tight timelines given depends on the quick and efficient submission of all collected data. Therefore, one single data submission format is proposed, and countries are asked to submit only one database to the international contractors. Furthermore, to avoid potential errors when consolidating the national databases, any changes in format that were implemented subsequent to the general agreement have to be announced.

Standard 12.1   Each PISA-D participant country submits its data in a single complete database, unless otherwise *agreed upon.*

Standard 12.2   All *data* collected for PISA-D will be imported into a national database using the Data Management Expert (DME) data integration software provided by the international contractors following specifications in the corresponding operational manuals and international/national record layouts (codebooks). Data are submitted in the DME format.

Standard 12.3   Data for all instruments are submitted. This includes the assessment data, questionnaires data, and tracking data as described in the PISA-D operations manuals.

Standard 12.4   Unless agreed upon, all data are submitted without recoding any of the original response variables.

Standard 12.5   Each PISA-D participating country's database is submitted with full documentation as specified in the PISA operations manuals.

Management standards

13. Communication with the international contractors

Rationale: Given the tight schedule of the project, delays in communication between the National Centres and the international contractors should be minimised. Therefore, National Centres need continuous access to the resources provided by the international contractors.

Standard 13.1 The international contractors ensure that qualified staff are available to respond to requests by the National Centres during all stages of the project. The qualified staff:

- Are authorized to respond to National Centre queries,

- Acknowledge receipt of National Centre queries within one working day,

- Respond to other queries from National Centres within five working days, or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.

Standard 13.2 The National Centre ensures that qualified staff are available to respond to requests by the international contractors during all stages of the project. The qualified staff:

- Are authorised to respond to queries,

- Are able to communicate in English,

- Acknowledge receipt of queries within one working day,

- Respond to queries from international contractors within five working days, or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.

Note 13.1 Response timelines and feedback schedules for the National Centres and the international contractor are further specified in the PISA operations manuals.

14. Schedule for submission of materials

Rationale: To meet the requirements of the work programme, and to progress according to the timelines of the project, the international contractor will need to receive a number of materials on time.

Standard 14.1 An *agreed upon Translation Plan* will be negotiated between each National Centre and the international contractors.

Standard 14.2 The *following* items are submitted to the international contractors in accordance with *agreed timelines*:

- the Translation Plan

- results from the national checking of adapted computer-based and paper-based assessment materials and questionnaires,

- adaptations to manuals,

- sampling forms (see Standard 1),

- demographic tables,

- other documents as specified in the PISA-D operations manuals.

Standard 14.3 *Questionnaire* materials are submitted for linguistic verification only after all structural adaptations have been *agreed upon*.

> **Standard 14.4**   All adaptations to those elements of the manuals that are required to be functionally equivalent to the source as specified in Standard 5.2, need to be *agreed upon.*

Quality Control

Agreed upon Translation Plan

International Contractors' records

Assessment materials submitted for linguistic verification with corresponding adaptation spreadsheets filled in by the National Centre

15.   Management of data

Rationale: Consolidating and merging the national databases is a time-consuming and difficult task. To ensure the timely and efficient progress of the project, the international contractors need continuous access to national resources helping to rule out uncertainties and to resolve discrepancies. This standard aims to prevent substantial delays to the whole project which could result from a delay in processing the data of a small number of participating countries.

> **Standard 15.1**   National Centres must make a **National Data Manager (NDM)** available upon submission of the database. The **NDM** must have the following qualifications:
>
> - Authorisation to respond to international contractor data queries,
>
> - Availability for a three-month period immediately after the database is submitted unless otherwise agreed upon,
>
> - Proficiency in English,
>
> - Ability to respond to international contractor queries within three working days, and to resolve data discrepancies.
>
> **Standard 15.2**   The timeline for submission of national databases to the international contractors is within six weeks of the last day of testing for the Field Trial and within six weeks of the last day of testing for the Main Survey, unless otherwise *agreed upon.*
>
> **Standard 15.3**   National Centres execute data checking procedures as specified in the PISA-D operations manuals before submitting the database.

Note 15.1   Each participating country/economy will receive its own national micro-level PISA-D database (the "national database"), in electronic form.

Each participating country/economy has access to and can publish its own data after a date that is established by the OECD Secretariat, the International Advisory Group and the PISA Governing Board for the publication of the initial OECD publication of the survey results.

The OECD Secretariat will not release national data to other countries/economies until participating countries/economies have been given an opportunity to review and comment on their own national data and until the release of such data has been approved by the national authorities.

A deadline and procedures for withdrawing countries/economies' national data from the international micro-level PISA-D database (the "international database") will be decided upon by the PISA Governing Board. Countries/economies can withdraw data only prior to obtaining access to data from other countries/economies. Withdrawn data will not be made available to other countries/economies.

The PISA Governing Board will discuss with participating countries/economies whose data manifests technical anomalies as to whether the data concerned can be included in the Strand C database. The decision of the PISA Governing Board and the OECD Secretariat will be final. Participating countries/economies may, however, continue to use data that are excluded from the international database at the national level.

The international contractors will then compile the international database, which will comprise the complete set of national PISA-D databases, except those data elements that have been withdrawn by participating countries/economies or by the PISA Governing Board at the previous stage. The international database will remain confidential until the date on which the initial international OECD publication is released.

National data from all participating countries/economies represented in the international database will be made available to all participating countries/economies from the date on which the initial international OECD publication is released.

After release of the initial international OECD publication, the international database will be made publicly available on a cost-free basis, through the OECD Secretariat. The database may not be offered for sale.

The international database will form the basis for OECD indicator reports and publications.

The international contractors will have no ownership of instruments or data nor any rights of publication and will be subject to the confidentiality terms set in this agreement.

The OECD establishes rules to ensure adherence to the above procedure and to the continued confidentiality of the PISA data and materials until the agreed release dates. These include confidentiality agreements with all individuals that have access to the PISA material prior to its release.

As guardian of the process and producer of the international database, the OECD will hold copyright in the database and in all original material used to develop, or be included in, the PISA Field Trial and PISA Main Survey (among them the assessment materials and manuals) in any language and format.

Quality Control

International Contractors' records from communication, forms, or documents

16.   Archiving of materials

Rationale: The international contractors will maintain an electronic archive. This will provide an overview of all materials and will also ensure that the international contractors have the relevant materials available during data cleaning, when they are first required.

Standard 16.1   The international contractors will maintain a permanent electronic archive of all assessment materials, manuals and other materials as appropriate, including all national versions.

Standard 16.2   The National Project Manager must submit one copy of each of the following translated and adapted Main Survey materials to the international contractors:

- a hard copy of all administered paper-based Questionnaires, if applicable, and,

- electronic versions of the manuals.

Standard 16.3   Unless otherwise requested, National Centres will archive all Field Trial materials until the beginning of the Main Survey, and all Main Survey materials until the publication of the international report. Materials to be archived include:

- all respondents' questionnaires (PBA version if applicable) or drives containing completed computer-systems and all associated data,

- sampling forms, and

- all data submitted to the international contractors.

After completion of a survey the National Centre will transfer this archive to the international contractors who will compile the national archives from all participants and transfer them to OECD after completion of the Main Survey.

National involvement standards

17.   National feedback

Rationale: National feedback is important in maintaining the dynamic and collaborative nature of PISA-D. National feedback ensures that instruments achieve cross-national, cross-cultural and cross-linguistic validity. It also promotes the inclusion of the interests and involvement of national stakeholders.

> **Standard 17.1**   National Centres should develop mechanisms in order to promote participation, effective implementation, and dissemination of results amongst all relevant national stakeholders.
>
> **Standard 17.2**   National Centres provide feedback to the international contractors on specific areas as necessary that represent the perspectives of the relevant national stakeholders.

Note 17.1   As a guideline feedback might be sought from the following relevant stakeholders: policy makers, curriculum developers, domain experts, test developers, linguistic experts and experienced teachers.

Quality Control

National Centre Quality Monitoring

Documented strategies

List of committees and groups

Membership records of representative groups and/or committees

Meeting records of representative groups and/or committees

18.   Meeting attendance

Rationale: Attendance at National Project Managers and training meetings is required as these represent a key component of participating in PISA-D. Important information is shared and discussed and training in data management, sampling, computer systems, and coding is conducted at these international meetings. These also allow for individual consultation and communication with contractors, which is often very helpful.

> **Standard 18.1**   Representatives from each National Centre are required to attend all PISA-D international meetings including National Project Manager meetings and coder training.
>
> **Standard 18.2**   Representatives from each National Centre who attend international meetings must be able to work and communicate in English.

Note 18.1   The length of these meetings usually varies from 3 to 5 days.

Quality Control

Meeting attendance records

Definitions

Agreed procedures - procedures that are specified in the PISA-D operations manuals, or variations that are agreed upon between the National Project Manager and the international contractors.

Agreed timelines - timelines that are specified in the PISA-D operations manuals, or variations that are agreed upon between the National Project Manager and the international contractors.

Agreed upon - variations and definitions agreed upon between the National Project Manager and the international contractors.

Basic dwelling -- A basic dwelling is a housing unit that has some but not all of the essential features of a conventional dwelling. It is a permanent structure or a part of a permanent structure, hence it may be a room or a suite of rooms in a permanent building but it is without some of the conventional dwelling facilities such as a kitchen, fixed bath or shower, piped water or toilet. In a number of countries or areas, a certain proportion of the housing inventory comprises such housing units which possess some but not all the characteristics of conventional dwellings.

Country of usual residence -- The country of usual residence is the country in which an individual lives, that is to say, the country in which he or she has a place to live where he or she normally spends the daily period of rest.

Temporary travel abroad for purposes of recreation, holidays, visits to friends or relatives, business, medical treatment or religious pilgrimage does not change an individual's country of usual residence.

Housing unit -- A housing unit is a separate and independent place of abode intended for habitation by a single household, or one not intended for habitation but occupied as living quarters by a household at the time of the census. Thus it may be an occupied or vacant dwelling, an occupied mobile or improvised housing unit or any other place occupied as living quarters by a household at the time of the census. This category includes housing of various levels of permanency and acceptability.

International Contractors' website – The PISA-D SharePoint can be accessed by authorised individuals through the following address: https://etsorg1.sharepoint.com/sites/ext/pisa-d/. This secure site contains the source versions of instruments, manuals and other documents and information relating to National Centres.

Link-tracing network samples – chain-referral samples used to survey hidden populations (hard to locate). The hidden population is identified by beginning with seeds (units belonging to the target population), and asking for referrals to other members of the hidden population. The referred to members are contacted for the survey, and also asked for referrals to others in the population. Limitations are that there is no good way to evaluate the coverage of the frame constructed in this manner, and duplications in the frame are not always evident. Examples of other chain-referral approaches are respondent-driven sampling (RDS) (Heckathorn, 2007) and snowball sampling (Goodman, 1961).

PISA-D Strand C Defined Target Population - all 14 to 16 year-olds in-school at grades 6 and lower, and 14 to 16 year olds who are out-of-school.

PISA-D Strand C Eligible Participants - Participants who are in the PISA-D Strand C Target Population.

PISA-D Operations Manuals - manuals provided by the international contractors, that is the following:

- Interviewer Procedures Manual,

- Data Management Manual,

- All other key documents referenced within the National Project Manager's manual.

The preparation of the PISA operations manuals will be carried out by the international contractors and will describe procedures developed by the international contractors. The manuals will be prepared following consultation with participating countries/economies, the OECD Secretariat, the Technical Advisory Group and other stakeholders.

Quota sampling - a non-random approach in which data collection stops when a target sample size is reached, even if all sample units have not been fully worked. Therefore, some bias is introduced into the sample because not all sample units were given a chance of selection.

Random route – an approach used for selecting households. The general approach is to have a master sample of starting points for each cluster. Then households are selected in the field using a pre-specified list (the 5th household, the 15th household, etc.) and a pre-defined route. As discussed in Murray, Kirsch and Jenkins (1998), for the random route household selection approach, selection probabilities are not generally known, and therefore the resulting samples are not considered probability samples. Furthermore, random route designs require control of sample selection in the field, and there is concern that the interviewers exert influence on the household selections. Control of the selection process by each country requires detailed instructions for selecting respondents that can be dependably followed by the interviewers. The random route approach differs from a systematic sample in which households are selected from a sampling frame using a specified sampling interval, and the interviewers are given exact addresses to attempt. Systematic samples are probability samples and are allowed for PISA-D Strand C.

Random seed - a number used to initialize a generator of random sequences (pseudorandom number generator).

Source Versions – the primary language of PISA-D documents is English. Some implementation documents will be provided in French and Spanish.

Study Management System – An ideally automated system that enables timely management, conduct and monitoring of data collection. The system facilitates the administration of the automated instruments, produces reports that are integral to the review and management of the data collection effort, and assists regional supervisors in their day-to-day operational tasks, such as case assignment and transfer and assignment of result codes.

Interviewer – an individual who is trained by the National Centre to administer the PISA-D Strand C instruments in household or school settings. .

Field Period - the period of time during which data is collected.

Translation Plan – documentation of all the processes that are intended to be used for all activities related to translation and languages.

REFERENCES

Bethlehem, J., Cobben, F., and Schouten, B. (2011), Handbook of nonresponse in household surveys. Wiley.

Biemer, P. and Lyberg, L. (2003). Introduction to survey quality. Wiley.

Carr-Hill, R. (2015), "PISA for development technical strand c: Incorporating out-of-school 15- year-olds in the assessment", *OECD Education Working Papers*, No. 120, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/5js0bsln9mg2-en

Couper, M., Singer, E., and Groves, R. (2008). Risk of disclosure, perceptions of risk, and concerns about privacy and confidentiality as factors in survey participation. *Journal of Official Statistics,* 24(2): 255-275.

Curtin, R., S. Presser and E. Singer (2000), "The Effects of Response Rate Changes on the Index of Consumer Sentiment," *Public Opinion Quarterly,* 64**:**413-428.

Goodman, L. (1961). Snowball Sampling. *Ann. Math. Statist.* 32, 148-170.

Groves, R. (2006), "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly,* 70:646-675.

Groves, R. and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. Public Opinion Quarterly, 72, 2, 167-189.

Heckathorn, D. (2007). Extensions of Respondent Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment. *Social. Methodology*, 37, 151-207.

Keeter, S., C. Miller, A. Kohut, R. Groves, and S. Presser (2000), "Consequences of Reducing Nonresponse in a National Telephone Survey," *Public Opinion Quarterly,* 64**:**125-148.

Mohadjer, L., M. Berlin, S. Rieger, J. Waksberg, D. Rock, K. Yamamoto, I. Kirsch and A. Kolstad (1997), "The Role of Incentives in Literacy Survey Research," in *Adult Basic Skills: Innovations in Measurement and Policy Analysis,* Hampton Press, Cresskill, NJ.

Mohler, P. and Johnson, T. (2010). Equivalence, Comparability and methodological Progress. In Harkness et al Survey Methods in Multinational, Multiregional and Multicultural Contexts. Wiley.

Murray, T.S., I. Kirsch and L. Jenkins (1998), *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey,* NCES 98-053, U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC.

Singer, E. (2002), "The Use of Incentives to Reduce Nonresponse in Household Surveys," in R.M. Groves *et al.* (eds.), *Survey Nonresponse*, Wiley, New York.

Stoop, I., Billiet, J., Koch, A., and Fitzgerald, R. (2010). Improving survey response. Wiley.

Tourangeau, R., Edwards, B., Johnson, T., Wolter, K., and Bates, N. (2014). Hard-to-sample Populations. Cambridge.

University of Michigan (2016). The comparative survey design and implementation initiative guidelines. https://www.csdiworkshop.org/index.php/initiatives/csdi-guidelines