

## Chapter 5

### Implementation procedures and approaches to including out-of-school children in educational assessments

*This chapter looks at two aspects of large-scale assessments. First, the implementation procedures used by PISA and other assessments, in particular institutional arrangements, sampling methods, data collection, data processing, and standardisation of implementation. In the case of each of the reviewed assessments, the chapter highlights any lessons that may be relevant for PISA for Development (PISA-D). Second, the chapter examines the methods and approaches that some of the reviewed surveys use to include out-of-school children in the assessments and highlights some lessons for PISA-D in this regard.*

## Implementation procedures

This section discusses the ways the reviewed assessments are implemented, with a view to highlighting any lessons that may be valuable for PISA-D.

Implementation is covered with reference to three different aspects:

- institutional arrangements
- survey implementation (including sampling, and data collection and processing)
- standardisation of survey implementation (including quality assurance).

### *International institutional arrangements*

Different assessments employ different structures to support the implementation of the assessment. In PISA, the PGB is made up of representatives from OECD member countries and countries that have associate status. Generally, representatives from countries that are not OECD members and do not have associate status attend PGB meetings as observers, although countries with long-standing experience in PISA can apply for full membership. The PGB determines PISA's policy priorities and ensures that these priorities are respected during each implementation.

For IEA studies there is a General Assembly, which provides overall direction for the development and implementation of the assessments. There are also meetings of national research co-ordinators, who guide the operational processes of the assessment.

SACMEQ is run by a consortium of education ministers from the participating countries. The consortium works with the UNESCO International Institute for Educational Planning (IIEP). The IIEP helps the countries design and implement the assessment and analyse the data it produces.

TERCE is organised and co-ordinated LLECE, a network of national education quality assessment directors across Latin America and the Caribbean. LLECE contributes to the global actions adopted by the UNESCO Regional Bureau of Education for Latin America and the Caribbean. These actions aim to ensure the right of all students in the region to have access to quality education. The TERCE pilot test was carried out in 2012 and the main study in 2013. All stakeholders worked together to develop the research tools and training to allow capacity building and the correct use of data.

Generally, the international institutional arrangements for the reviewed large-scale international assessments involve a governing group or steering committee to set overarching policies and priorities, and one or more groups to provide technical guidance.

The household-based surveys also follow this model. One key difference between these surveys is the extent to which external organisations as a whole are enlisted to assist with the more technical aspects of design and implementation. STEP and LAMP both made use of external organisations as a whole (the Educational Testing Service in STEP's case, and the Educational Testing Service and Statistics Canada for LAMP). ASER and Uwezo prefer to draw on networks of experts and consultants.

For PISA-D, its international institution arrangements will include having observer status on the PGB (OECD, 2014a: 26). PISA-D may also receive advice from other PISA-affiliated boards and groups.

The other international groups in PISA are the Technical Advisory Group, the Subject Matter Expert Groups (SMEGs – one for each cognitive assessment domain), and the Questionnaire Expert Group. Members for these groups are nominated by the international contractors for PISA. The Technical Advisory Group advises on all technical matters, and the Subject Matter Expert Groups and Questionnaire Expert Group prepare the theoretical frameworks.

The existing PISA Technical Advisory Group, with the addition of experts in assessment in developing contexts and in the area of out-of-school youths, will serve as the body that gives technical advice for PISA-D (OECD, 2014a: 28-29).

The Subject Matter Expert Groups and Questionnaire Expert Group for PISA will support PISA-D in a framework and instrument development process facilitated by the OECD Secretariat and the international contractor for the project (OECD, 2014a: 28).

The new international body established to particularly serve PISA-D is called the International Advisory Group. It will be made up of representatives from participating countries, development partners, institutional partners and the OECD. Some experts will also be part of this group. The International Advisory Group's responsibilities reflect its varied membership. The group will advise the OECD and the PGB on priorities for PISA-D. It will give input into the selection of international contractors for the project and guide project implementation. By representing the interests of the different constituent stakeholders, the International Advisory Group will contribute substantive expertise to the project. It will also be responsible for assessing progress and reviewing expenditure (OECD, 2014a: 28-29).

It would appear that none of the international institutional arrangements for the reviewed assessments include a group with a brief as varied as PISA-D's International Advisory Group. It is possible that the group's role may be made even more complex by the fact that a number of separate one-to-one relationships might exist between its different members – individual relationships between development partners and participating countries, for example.

Building capacity in participant countries, and peer-to-peer learning with non-OECD countries that participate in PISA are described as key features of PISA-D (OECD, 2014a: 61). These features will most likely influence the institutional arrangements at the international level.

The international contractor for PISA-D and the International Advisory Group will need to facilitate capacity-building activities. This may require the formation of additional groups or partnerships, perhaps even across PISA and PISA-D boundaries.

### ***National institutional arrangements***

In PISA, each participating country establishes a national centre, usually within a government department, university or other research institution. PISA-D is adopting the same model. In addition to establishing a national centre led by a national project manager, PISA-D participating countries will be expected to convene a national committee of experts in assessment and individuals drawn from the education and scientific communities. In each participating country, the national committee will be expected to advise the country representative on the International Advisory Group and the national project manager about the suitability of the data collection instruments, quality control issues and national reporting and dissemination.

All of the reviewed assessments that are undertaken in multiple countries also have a national-level group in their institutional arrangements. There are few instances of an equivalent to the proposed PISA-D national centre. Instead, in each country there is just one national-level group, and this group communicates directly with the country's representative in the international-level groups.

In several of the reviewed assessments, the national-level division of activities and responsibilities varied from one participating country to the next. For example, in WEI-SPS, in some countries survey operations were outsourced in their entirety to private companies; in other countries staff from ministries or other educational institutions conducted survey operations; and in other countries again the national-level implementation centre hired and trained personnel to undertake survey operations (UIS, 2009a: 55). PISA-D may need to be prepared to accommodate a similar variety of national-level arrangements. There was also some variation in national-level arrangements in LAMP. The LAMP evaluation states that making the different national-level partnerships operational was one of the challenges faced by the national project teams (UIS, 2008).

As mentioned above, the call for tender for the international contractor of Strand A and Strand B of PISA-D emphasises the capacity-building aspect of the project. At the national level, the results of the capacity needs analysis and the capacity-building plan will most likely inform the establishment and scope of work of the national centres.

### ***Implications***

None of the reviewed surveys included an overall governing body whose composition and described role is as broad and varied as that envisaged for the PISA-D International Advisory Group. Careful consideration will be needed so that the International Advisory Group can fulfil its broad and varied role and accommodate the interests from the distinctly different stakeholder groups of which it is composed. For example, meetings of subgroups of the International Advisory Group may need to be scheduled to address different stakeholder interests.

PISA-D will need to consider how to best formalise the capacity-building and peer-to-peer learning emphases in institutional arrangements at the international and national levels. For example, partnerships could be established between PISA-D participating countries and PISA countries that have similar capacity needs. PISA-D countries should be encouraged to establish their national centres with a view to getting the most value out of any capacity-building support that is provided.

The roles and responsibilities of the national committee will need to be clearly described, and each participating country given guidance, to ensure that a productive relationship is established between the national committee and the national centre.

In several of the reviewed assessments, the national-level division of activities and responsibilities varied from one participating country to the next. A LAMP evaluation stated that making the different national-level partnerships operational was one of the challenges that countries faced in their implementation. There are likely to be a variety of management arrangements for PISA-D. These might range from full outsourcing of some activities to national centre staff handling everything. PISA-D should consider: how to support national centres to manage in-country relationships; and how to effectively communicate quality assurance requirements so all involved parties understand them.

## Survey implementation

### *Sampling*

#### *Target population definition*

The definition of target population varies across the international assessments reviewed, depending on the priorities of the assessment as expressed through assessment frameworks. Once the level of education is decided (for example lower secondary, upper primary, lower primary) the population then needs to be defined operationally. The main decision is whether to sample students by age or by grade. Apart from PISA, all the large-scale international surveys reviewed define target populations by grade, with the lowest grade being Grade 2 for PASEC, and the highest grade being Grade 8 for TIMSS. A grade-based population definition is thought to be more appropriate for these surveys, because they are strongly curriculum referenced.

At the same time, the household-based surveys define target populations by age, with the narrowest age range being 6 to 16 years old for Uwezo, and the widest age range being 15 years old and older for LAMP.

PISA aims to test children approaching the end of compulsory schooling. The target population definition is 15-year-old students in Grade 7 or higher who attend educational institutions full-time or part-time, vocational training programmes or other related types of educational programmes, or foreign schools.

PISA includes some countries in which school is not compulsory for 15-year-olds. In some of these countries (such as Croatia and Korea), many 15-year-old children who have not continued in traditional schooling beyond the age at which it ceases to be compulsory can be covered by the PISA target population definition through its interpretation of “school” as including institutions that offer part-time education and vocational education programmes. In other countries (such as Brazil and Mexico), many 15-year-olds are not covered by the definition because they did not continue with any form of education beyond the age of compulsory schooling.

PISA-D will use the same target population definition as PISA. Across the six PISA-D participating countries, school is compulsory until the age of 16 in Senegal; 15 in Cambodia; and 14 in Ecuador, Guatemala, Paraguay and Zambia (OECD, 2014a: 18).

The situation in the PISA-D pilot countries may be similar to that in Brazil and Mexico, in that beyond the age of compulsory schooling there is a notable number of children who are not receiving any kind of education, whether full-time, part-time or vocational. Some of these children may, of course, be covered by PISA-D’s attempts to sample out-of-school children.

An age-based target population definition in PISA-D will have to take into account the grade distribution of 15-year-old children and the age distribution of children in the “national modal grade” for 15-year-olds, both within one participating country and across all participating countries. The national modal grade is the year level attended by most 15-year-olds in a country.

In PISA-D, participating countries may choose to assess 15-year-old students in grades below Grade 7 and out-of-school children, but if countries do choose to assess these populations, the results will not be included in the national results that are reported on the PISA scales – the aim will be to develop an approach and methodology for

assessing these populations rather than to obtain nationally representative results for them (OECD, 2014a: 51). Methods and approaches to include out-of-school children are discussed in section 5.2 of this report.

### *Sampling approach*

PISA employs a two-stage stratified sampling methodology. Schools are sampled first, and within-school units – either classes or students – are sampled second. School exclusions and within-school exclusions are permitted for political, organisational or operational reasons, as long as they do not exceed set limits.

PISA offers flexibility in sampling options so participating countries can investigate in more depth the performance of particular subpopulations of interest. This may be of interest to PISA-D participant countries, who may want to test children who are not 15 years old but who are in the national modal grade for 15-year-olds, if there is a significant number of them.

In PISA, schools are sampled centrally by an organisation contracted by the OECD. Within-school sampling is undertaken by the national centres using software specifically developed for this purpose. To achieve this second-stage sampling, sampled schools are required to submit to accurate and complete lists of children who are eligible to be sampled.

All the reviewed surveys employ a multi-stage sampling methodology. Of the large-scale international surveys:

- SACMEQ samples schools, then sample subsets of children across all classes in the target grades of sampled schools.
- PASEC samples schools, then one classroom (of each grade) is selected among the available classrooms of a particular Grade (2 or 6), and finally students are selected from the list of all students of the classroom.
- LLECE has in the past sampled all children in the target grades in sampled schools, but for TERCE, sampled one classroom per school per grade.
- The IEA studies PIRLS, prePIRLS and TIMSS sample schools, then samples intact classes from the target grades in sampled schools.
- Of the other school-based surveys and the household-based surveys:
  - EGRA and EGMA sample schools, and sample subsets of students from sampled schools
  - the household-based surveys sample households, and then sample from individuals within the target population in the sampled households.

It is unclear from the call for tender for the international contractor for PISA-D whether this project will aim to adhere to all of the PISA technical standards that relate to sampling. If PISA-D intends to follow all these technical standards, several questions will need to be addressed for all current and potential PISA-D participating countries:

- Will security issues mean that in some instances a desired target population is not able to cover the entire country?
- Will up-to-date and complete lists of schools be available to form the basis of the school sampling frame?

- Will schools be able to supply complete and up-to-date student lists so the national centres can draw the student samples?

With respect to the impact of security issues on target populations, it's possible that a country may only be accepted to participate in PISA-D if its security situation is considered stable. This may mean that interested countries with unstable areas cannot be involved. At the same time, participation in PISA is a multi-year process, and situations can change quite dramatically over such a period of time.

With respect to the currency and completeness of lists of schools ASER provides a relevant example. While ASER is a household-based assessment, it also includes a component where test administrators visit schools to record information about facilities and resources and to observe a class. The test administrator is instructed to visit the school attended by most of the children in the village. One of the reasons that schools are not sampled before the testing day is because a complete, up-to-date and official list of schools is not available in India. In particular, many low-cost private schools are not recognised by government.

SACMEQ provides another example of dealing with the issue of currency and completeness of school lists. In SAQMEQ, children are sampled on the day of the test by data collectors (SACMEQ, 2007a). This means that the test administrators can double-check the list on the testing day, but it does place an additional burden on the data collectors.

One of the main reasons ASER tests children in households rather than in schools is because it aims to test all children within the target age range, and testing in school is not the way to access a sample of all children; some children have dropped out of school, others attend but only irregularly, and some have never been enrolled (ASER Centre, 2014). If the situation in some PISA-D countries is similar to that described by ASER in India, the OECD may need to consider the terminology and phrasing it uses to discuss the results from its school-based assessment. Implying that the results give a picture of what all 15-year-olds can do may not be appropriate.

### ***Data collection***

In PISA, the cognitive assessments are paper-based and computer-based. They are administered in schools to groups of children, but each child completes the assessment independently. The questionnaires are also paper-based and computer-based. PISA is in the process of transitioning to completely computer-based data collection. The call for tender for the international contractor for Strand A and Strand B of PISA-D states that in this project the instruments will be paper-based (OECD, 2014a: 32).

In terms of how cognitive data are collected, the reviewed surveys can be broadly categorised:

- In several surveys, the cognitive assessment is a paper-based instrument that is administered in schools to groups of children, and each respondent completes the assessment independently by reading questions and recording responses on paper. Surveys that fit this category are PIRLS and prePIRLS, TIMSS, LLECE, SACMEQ and PASEC Grade 6.
- In other surveys, the cognitive assessment is a paper-based or computer-based instrument that is administered one-on-one, either in households or in schools. Within this category, there are:

- surveys in which the respondent completes the cognitive assessment independently of the data collector, by independently reading questions and recording responses on paper or entering them into the computer; LAMP, PIAAC and STEP fit this subcategory
- surveys in which the data collector delivers the cognitive assessment orally and the respondent gives most, if not all, of his or her answers orally. EGRA, EGMA, ASER, Uwezo and PASEC Grade 2 fit this subcategory. In some administrations of EGRA and EGMA, the student has a paper-based test, but the test administrator collects data using a tablet-based application called Tangerine™.<sup>1</sup>

These different administration methods for cognitive assessments suit different survey aims and purposes. Group administration is most convenient if members of the target population are expected to be proficient enough to complete an assessment independently, and easily located in naturally occurring groups (such as in schools). Household-based administration may be required if naturally occurring groups of members of the target population cannot be effectively accessed to facilitate group administration. One-on-one oral administration is necessary if some respondents are expected not to be proficient enough to complete an assessment independently. If one-on-one oral administration is used, a tablet-based data collection application such as Tangerine™ is an effective way to reduce recording errors.

In terms of how contextual data are collected, the reviewed surveys can be broadly categorised as:

- surveys in which respondents are asked to complete questionnaires, including PIRLS and prePIRLS, TIMSS, LLECE, SACMEQ, WEI-SPS and PASEC Grade 6; and
- surveys in which respondents are interviewed by data collectors, including EGRA, EGMA, STEP, LAMP, ASER, Uwezo and PASEC Grade 2.

Contextual data collection by interview is less cost-efficient since interviews must be conducted one-on-one, but it is generally considered to lead to lower incidences of missing data and non-response.

With respect to data collection overall, whether this process involves the use of computers or paper-based instruments has implications beyond the test administration. If computers are used, then the subsequent data capture step is not required (as discussed further below).

The process for sourcing individuals to be involved in data collection varies across the reviewed surveys. In the IEA surveys prePIRLS, PIRLS and TIMSS, the sampled schools or the national research co-ordinators appoint school co-ordinators, and the school co-ordinators identify suitable test administrators (Johansone, 2012). In the other reviewed surveys, our understanding is that selecting test administrators is the responsibility of the national centres, their regional delegates, or the local contractor responsible for data collection. An important aspect of ASER and Uwezo is the use of local volunteers for test administration. It has been found that volunteers carry out the tasks assigned to them very effectively (Results for Development, 2015). Sourcing adequate numbers of volunteers is facilitated by partnerships between the national ASER and Uwezo offices and local organisations.<sup>2</sup>

There are some particular elements of the data collection approaches of SACMEQ and LLECE that may be instructive for PISA-D.

In both surveys, data collection in a school happens over a number of days, with the cognitive and contextual sessions on different days (LLECE, 2010: 37; SACMEQ, 2007b: 58).

SACMEQ III had a student homework form, containing questions that were previously included in the student questionnaire but that children might be better able to answer with the help of family members. This form included questions about parental level of education, home possessions, time taken to get to school, whether or not biological parents were alive and so on. Students took this form home on the first night of the survey administration and were expected to bring it back the next day. This method considerably reduced the number of missing values in the SACMEQ III study compared with previous SACMEQ studies (Hungu, 2011a: 4).

In LLECE, test administrators have a suggested time for the cognitive sessions, but they are permitted to allow up to ten minutes of additional time if necessary (LLECE, 2010: 37). However, this is no longer the case for TERCE. In SACMEQ, data collectors are also given a suggested time for the testing sessions. They are not obliged to keep to it, however, and it is stated that they can reasonably allow an additional 50% of the actual session time as extra time (SACMEQ, 2007a, 2007b: 57).

At the end of each cognitive session and contextual session, SACMEQ test administrators check each student booklet, and if they see any items that have not been completed, they ask the student to complete them (SACMEQ, 2007a: 23).

In SACMEQ I, Zambia had significant non-response rates, for reasons that were not always clear from the records submitted by the test administrators. In the Luapula region, for example, 4 of the 15 schools selected in the sample either refused to participate or were not visited by the data collectors. Further non-response occurred in Luapula because 10% of the pupils in the remaining 11 schools were absent on the day of testing (Nkamba and Kanyika, 1998). In SACMEQ III there were major data losses due to the loss of data collection instruments. In the end, information was only obtained from 61% of selected schools (Musonda and Kaba, 2011). These rates of non-response and data loss were not reflected in other SACMEQ countries.

## ***Data processing***

### *Coding*

In PISA-D, coding for constructed response items will be undertaken within the participating countries. Different procedures will ensure coding quality, including coding verification by expert coders and a coder reliability study across all participating countries (OECD, 2014a: 43).

PISA is not only careful to ensure coding quality for cognitive data, but also for occupation data. In past PISA cycles, countries have been given the option of incorporating double coding of occupation data in their internal training processes.

In regard to coding, the reviewed surveys appear to expend considerable time and resources on coding training and coding itself, including the steps taken to confirm that coding is being undertaken with acceptable reliability.

In PIRLS, prePIRLS and TIMSS, comprehensive coder training is provided. Responses to use in coder training are obtained from real children. In prePIRLS, the responses were obtained from children in Botswana, South Africa and the United States of America. The IEA studies use a qualification database that includes responses from previous cycles of the assessment and the codes assigned to those responses. The current year's coders are required to code these responses from previous years. They cannot begin coding current year responses until a specified level of agreement with the codes assigned to responses from previous years has been achieved. Studies of coding reliability within the current year and across previous years are conducted in each participating country. Cross-country reliability studies are also conducted. All of these studies use software developed specifically for the purpose by IEA's Data Processing and Research Centre.

In LLECE, coder training is provided centrally to national representatives, who are then expected to replicate the training with their national coding teams in their own countries. Responses to constructed response items are coded twice.

In EGRA and EGMA, coding is undertaken at the time of test administration, and coding training forms part of test administrator training. Organisations implementing EGRA and EGMA are advised to investigate and improve inter-rate reliability during the training (RTI International and International Rescue Committee, 2011).

In PIAAC, participating countries that used a paper-based assessment were required to undertake in-country reliability studies in both the field trial and the main survey. In these studies, a predefined number of responses were coded by a second coder, and the level of agreement had to be at least 95%. Cross-country reliability studies were also conducted to identify any systematic coding bias across countries. In these studies, bilingual coders coded responses in source booklets in English and in the national language booklets. Codes were compared across languages and also to the scores assigned to the same responses by master coders (OECD, 2013a: 59).

STEP employs similar systems as PIAAC for training in coding and for coding evaluation (Pierre et al., 2014: 61).

In LAMP, the UIS trained a chief coder for each country, and this coder trained all national coders (UIS, 2004: 38). UIS also specified amounts of material that had to be double coded within countries. One hundred per cent of tests had to be double coded during the field trial. In addition, a sample of responses from each country was coded by an international coder to ensure no bias at the country level (UIS, 2009b: 40).

In PASEC, this operation is performed by a group of coder and data entry clerks that are recruited and trained by the national team with support from PASEC technical advisors in the country. The technical advisor presents the patterns of the responses for each instrument to the data entry clerks. A manual of all coding instructions for all multiple choice questions is given to the data entry clerks.

The manual for the national research co-ordinator for SACMEQ III (see SACMEQ, 2007b) does not mention coding at all, which suggests that this survey has opted to not include test or questionnaire items that require human coding.

### *Data capture*

In regard to data capture, if the reviewed surveys use paper-based administration, human data entry is undertaken at the country level using a standard application

developed specifically for the purpose. For surveys that use computer-based administration, there is no separate data capture step following test administration. One exception to this is EGRA and EGMA, in which the test instrument is paper-based, but in many administrations the data collectors record children's responses using a tablet-based application (as mentioned above).

For the surveys that use human data entry, there is some variation among types of data entry software. PIRLS, prePIRLS and TIMSS all use IEA's own WinDEM software application (Johansone, 2012: 14-15). In WEI-SPS, countries were also encouraged to use WinDEM (UIS, 2009a: 63). LLECE and ASER both use Access-based data entry applications (LLECE, 2010: 45; W. Wadhwa, personal communication, 8 August 2014). Uwezo uses Stata (Uwezo, 2013a: 1). The guidelines for planning and implementing EGRA suggest programmes such as Excel, Access, CS Pro and FileMaker can be used for data entry. These guidelines recommend that implementing organisations consider the balance between, on the one hand, giving local personnel longer-lasting and more applicable skills, which may be better achieved with simpler data entry processes developed in common software such as Excel; and, on the other hand, applying an adequate level of rigour to data entry, which may be better achieved using a more complex software but one that local personnel might have difficulty using and might never be required to use again (RTI International and International Rescue Committee, 2011: 62-63).

None of the surveys that make use of paper-based data collection appear to use scanning rather than human data entry.

All of the reviewed large-scale international surveys require some percentage of data to be entered twice as a means of verifying the quality of data entry. With respect to this data verification step, in PIRLS, participating countries and benchmarking entities are required to double-enter 5% of their data, but in the case of South Africa's participation in PIRLS 2011, this amount was increased to 100%, based on experience in previous studies (Howie et al., 2012: 26). SACMEQ requires 100% of data from tests and questionnaires to be entered twice (SACMEQ, 2007b: 63).

The call for tender for the international contractor for Strand A and Strand B of PISA-D suggests that this project will follow a similar model to PISA for data capture. That is, this activity will be the responsibility of in-country teams, but they will use software developed by the international contractor and they will be trained by the international contractor to use this software (OECD, 2014a: 44). The call for tender does not mention whether there will be any requirements for double data entry in PISA-D.

### *Data cleaning*

If PISA-D follows the same model as PISA, participating countries will complete some initial data validation before submitting their data to the international contractor. The international contractor will then undertake data cleaning according to standardised procedures. In PISA, data cleaning is a comprehensive iterative process that can involve several weeks of backwards and forwards communication between the international contractor and the national centre of a participating country.

In regard to data validation and cleaning, from the available information it appears that for many of the reviewed surveys these activities also follow a similar model to PISA. In this model, preliminary data validation is undertaken at the country level, and more comprehensive data cleaning is undertaken centrally.

For a number of the reviewed surveys that use paper-based data collection, data validation checks are built into the standardised data entry application. This step actually begins before human data entry, when data collection sheets are checked for completeness and correctness at the time of or immediately after administration.

In SACMEQ, data collectors are expected to spend some time checking all test booklets and questionnaires for missing and discrepant data, and make attempts to rectify any issues before leaving the school. The SACMEQ III data collector's manual describes the required checks. The full set of checks is quite considerable, and includes not only checks for missing data, invalid values and discrepant combinations of values within particular instruments, but also checks across different instruments (SACMEQ, 2007a: 43-50).

PASEC technical advisors build some macros to check the patterns of the responses, duplicates, the way the filter questions have been filled in, the expected values for the variables, and the concordance of students' participation between students tracking forms and tests data. This step usually leads to more data entry at the national level. A procedures manual has been developed and provides all the details of data cleaning. At this step, the sampling coverages are calculated, together with other indicators such as the weighted and non-weighted participation rates. PASEC uses the rules of PISA to classify countries regarding the reliability of the data collected. Replicate weights (a series of variables that contain the information necessary for correctly computing (via the replicate weight method) the standard errors of point estimates when analysing survey data) are also generated at the end of this process.

### *Implications*

We suggest that the OECD should consider subnational participation arrangements so that countries with stable and unstable areas might be able to participate in PISA-D.

Deliberation will be needed to work out how a school sampling frame that satisfies PISA's technical standards will be constructed in countries where complete and up-to-date lists of schools are not maintained. Of note is that among the surveys reviewed, ASER eschews school sampling because such complete and up-to-date lists of schools do not exist.

The OECD could consider whether PISA's approach to student sampling is appropriate in contexts where schools do not maintain complete and up-to-date lists of students. Notably among the surveys reviewed, SACMEQ has test administrators sample students on the day of the assessment, which gives them an opportunity to double-check the student list on the testing day.

In terms of data collection, we suggest that the OECD considers whether any of the following approaches may be appropriate to incorporate into PISA-D:

- interview sessions to collect contextual data from respondents other than students (such as principals and teachers), perhaps using a tablet-based data collection tool to eliminate recording errors
- cognitive test administration over multiple days (as done in LLECE and SACMEQ)
- permitting extra time to complete cognitive assessments (as done in LLECE and SACMEQ)

- establishing on-site test administrator checks of student booklets to reduce the incidence of missing data (as done in SACMEQ)
- sourcing test administrators who are local to test administration sites as a means of securing community engagement and buy-in (as done in ASER and Uwezo).

If the OECD intends to confirm coding reliability within and across PISA-D countries, it may be useful to follow the approach used by IEA, in which responses and their codes are pre-loaded into a database.

The platform and complexity of the data capture software that is provided to participating countries will be central to ensuring that data capture activity adequately serves the project's aims for sustainable capacity development. The guidelines for planning and implementing EGRA raise this issue (see RTI International and International Rescue Committee, 2011: 62-63).

We suggest considering more stringent requirements for double data entry than are currently implemented in PISA. SACMEQ requires all countries to double-enter 100% of test and questionnaire data. The most recent PIRLS cycle required South Africa to double-enter 100% of its data.

In relation to data cleaning, it may be advantageous for PISA-D to include data validation steps for test administrators to carry out before they leave the schools, as is done in SAQMEQ. Including these steps may simplify processes and reduce subsequent data cleaning activities.

## Standardising implementation

### *Articulation of standards*

PISA has a range of technical and operational standards that are articulated in a specific standards document. These standards cover aspects of implementation that have a direct impact on data quality, management standards that address operational objectives, and national involvement standards (OECD, n.d.: 4-5). PISA also produces comprehensive manuals to support countries to conduct survey operations in adherence to the standards.

In most of the reviewed surveys, standards are also typically articulated through specific standards documentation, or through the instructional materials prepared to guide implementation. The following details may be of particular interest to PISA-D.

The WEI-SPS technical report highlights the difficulty in establishing standardised procedures when the participating countries are geographically, culturally and economically diverse. The report states that the WEI-SPS standardised processes were refined after the pilot, and that countries were still able to deviate from them with approval (UIS, 2009a). This flexibility may be something that PISA-D will need to consider.

In LAMP, standards were articulated in the memorandum of understanding that each participating country signed with UIS (UIS, 2009b). In STEP, the standards are articulated in the National Survey Design Planning Report that each participating country was required to complete (for example, see World Bank, 2013). These documents are specific to each participating country rather than general to all participating countries. Incorporating at least some of the standards into a memorandum of understanding or

survey implementation plan may be an effective way to ensure that participating countries are fully aware of their responsibilities with respect to the standards.

Uwezo publishes a standards document online that covers survey implementation and also behavioural standards for the different roles in the initiative (see Uwezo, 2012). This document is a distillation not only of the “nuts and bolts” of the survey, but also of its underlying values.

### ***Training***

In PISA, national representatives are trained at international training sessions, and where necessary, these national representatives then train their in-country personnel. The call for tender for the international contractor for Strand A and Strand B of PISA-D suggests that PISA-D will follow the same approach as PISA.

With respect to training, all the reviewed surveys follow some kind of cascade training model similar to that used in PISA. In the case of the large-scale international surveys, the institution with overall responsibility for the survey trains national co-ordinators centrally. These national co-ordinators or administrators take on the role of trainers for others at the national level as required.

At the international level, training typically covers everything from coding and data management, to translation and adaptation (if applicable), test design and development, sampling, and analysing national data.

Angola’s Ministry of Education is participating in SACMEQ IV as an observer (SACMEQ, 2013). Representatives from this ministry attend international-level SACMEQ training.

SACMEQ and PASEC have delivered a number of workshops for participants (SACMEQ, 2015a). The International Institute for Educational Planning has also prepared training modules (SACMEQ, 2015b). These training modules are a useful general resource – new SACMEQ participants could use them to become familiar with the key aspects of survey design and implementation before they delve into the more specific details of the SACMEQ project.

In the reviewed surveys, the number of tiers of training at the national level varies. This depends on features of the survey, including the sample size (and therefore required number of data collectors), geographical scope and languages of test administration. In regard to data collection training in particular, the SACMEQ adopts a training approach in which regional research co-ordinators and team leaders are hired and trained centrally. These regional research co-ordinators then host training for data collectors within their respective regions (SACMEQ, 2007b). LLECE also has a broad capacity development programme.

The household-based surveys ASER and Uwezo also incorporate multiple levels of training below the national level. In ASER, the national ASER Centre staff are trained through a national workshop. These staff members go on to train master trainers at state workshops. The master trainers then train volunteer data collectors at district workshops (ASER Centre, 2014). In Uwezo, key facilitators are trained at the national level. Key facilitators are responsible for training master trainers, who themselves train district co-ordinators, who train volunteer data collectors (Uwezo, 2012). These two household-based assessments incorporate multiple levels of training below the national level for the reasons given above (sample size, geographical scope and languages of test

administration), but also because the surveys aim to always use local volunteers to collect data at each survey administration site.

In Uwezo, representatives from each of the three countries (Kenya, Tanzania and Uganda) observe and participate in cross-country training, as a way of ensuring that best practices are shared (Uwezo, 2012: 13).

For many of the reviewed surveys, data collectors and supervisors are given the opportunity – during training or during a field trial – to participate in in-the-field practice of the elements of survey administration in which they play a role. Elements might include, for example, sampling children or households, administering the survey or completing data collection sheets. Offering this practice in the field is an important step in ensuring the data collectors fully understand their responsibilities and have been adequately trained in executing them. Ideally, the data collectors should be supervised during this in-the-field practice and given an opportunity to debrief, so that any difficulties or issues that were encountered can be fully discussed.

In SACMEQ, the in-the field practice itself follows the cascade model: at the national training, experienced regional research co-ordinators demonstrate test administration at the trial schools, and team leaders and other regional research co-ordinators observe and take notes. At the regional training, experienced team leaders then demonstrate test administration and the data collectors observe and take notes. The SACMEQ regional training sessions typically occur in the week immediately before the main data collection (SACMEQ, 2007b: 52-53).

In most of the reviewed international large-scale surveys there are two levels of instructional documentation: one level for the national co-ordinator of the assessment, and another level for the data collectors. Both levels of documentation are generally prepared by the institution with overall responsibility for the survey and translated for national use if necessary.

The IEA PIRLS and TIMSS studies present national-level instructional material in units. For the surveys in 2011, there were seven separate units for sampling, field trialling, contacting schools and sampling classes, preparing materials for data collection, collecting data, scoring, and creating data files. The units are released at key points in the survey timeline. Some units were supplemented by other manuals (such as for test administrators) or IEA-developed software (Martin and Mullis, 2012). Releasing instructional materials in units may make the large quantity of information more easily digestible, while still enabling national-level staff to get a sense of the coherent whole.

The manual for the national research co-ordinator for SACMEQ III contains a comprehensive and easy-to-read timetable of activities. This timetable gives a clear overview of the full range of responsibilities at the national level (see SACMEQ, 2007b: 9-13).

Uwezo prepares manuals to support trainers of test administrators, as well as manuals and workbooks for test administrators themselves (see, for example, Uwezo, 2013b). Much like Uwezo's standards document (mentioned above), these materials serve an informational purpose but also reflect the survey's underlying values and ideology in an effective way.

PASEC is preparing a procedure manual for 2016.

### *Quality assurance*

The call for tender for the international contractor for Strand A and Strand B of PISA-D refers to the “stringent quality-assurance mechanisms that are applied to test design, translation, sampling and data collection” in PISA (OECD, 2014a: 18). Monitoring activities, outcomes and outputs against the articulated standards is a central part of this quality assurance.

Most of the reviewed surveys have at least some quality assurance mechanisms in place. The mechanisms related to translation and adaptation, coding, and data management have been discussed in earlier parts of this document. One other key aspect of quality assurance is the monitoring of the data collection activities.

In the reviewed surveys, the quality of the data collection activities is monitored via the following avenues:

- Selected survey administrations are monitored by international quality monitors appointed by the institution with overall responsibility for the survey – prePIRLS, PIRLS, TIMSS, LLECE.
- Selected survey administrations are monitored by national quality monitors associated with or employed by the national project team – prePIRLS, PIRLS, TIMSS, SACMEQ, LLECE, PASEC.
- Monitoring supervisors complete administration reports that highlight any observed deviations from standardised administration – prePIRLS, PIRLS, TIMSS, SACMEQ, LLECE, PASEC.
- Data are checked soon after administration, and re-survey is undertaken in instances where issues are identified – ASER, Uwezo.
- National project teams summarise information from monitoring supervisors and include the summary in documents submitted to the institution with overall responsibility for the survey – prePIRLS, PIRLS, TIMSS, SACMEQ, LLECE, PASEC.

The quality of data collection is most effectively monitored by individuals who are familiar with the survey and the context in which it operates, but who have some level of independence from the survey administration team.

If quality issues are identified during or soon after data collection through a rigorous and independent monitoring process, then the survey has the best possible opportunity to take the necessary remediation steps.

Another key aspect of quality assurance is maintaining the security of the test materials. This appears to be of particular concern for SACMEQ and PASEC. The manuals for the national research co-ordinator and the data collectors include numerous “warning boxes” reminding readers of the importance of maintaining the materials’ security. Steps to maintain this security include (SACMEQ, 2007a, 2007b):

- advising national centre staff that the test should not be distributed to printers when requesting quotes for printing
- encouraging regional research co-ordinators to collect their own bundles of materials from the national centre

- reminding the national research co-ordinators that it is ultimately their responsibility to track any packages that have been dispatched from their offices
- impressing on data collectors that after test administration all materials must be taken away from schools, and that if tests are left in the hands of teachers at schools, they will most likely be used in class
- impressing on data collectors that they cannot let anyone copy the test materials.

### *Implications*

In relation to standards articulation, we suggest that the PISA-D should carefully consider whether at least some standards should be articulated in a memorandum of understanding or project implementation plan as well as in a dedicated standards document. Including the standards in documents that are specific to each participating country, rather than general documents, may be effective as a means of ensuring that each country is fully aware of its responsibilities with respect to the standards. LAMP and STEP both articulate their standards within the national project implementation plans.

We also suggest that the OECD considers how the description of standards can be used as an opportunity to reflect PISA-D’s underlying values and ideology in a way that will help to secure local commitment to the project and acceptance of its results. Of note among the surveys reviewed, Uwezo does this effectively.

In relation to training, the reviewed large-scale international surveys accommodate a variety of country capacities within a standardised international-level training module. Generally, this information is not available in the public documentation for the surveys, so further questioning of the institutions responsible for the surveys may be required. PISA-D training processes will need to balance this standardised training with specific targeted training that is tailored based on the findings of the capacity needs analyses.

In relation to quality assurance, further information is required from SACMEQ, PASEC and LLECE about their quality assurance processes, particularly those related to assuring the quality of test administration. These surveys may be able to highlight common risks and pitfalls that will be instructive for PISA-D.

## **Methods and approaches to include out-of-school children**

PISA makes no attempt to assess out-of-school children. The summary record from the first meeting of PISA-D’s International Advisory Group states that PISA-D will not aim to complete a system-wide assessment of out-of-school children, but will “explore approaches to addressing out-of-school 15-year-olds, including building on existing initiatives and piloting sampling methods, test items and background questionnaires on smaller convenience samples” (OECD, 2014b: 5).

In preparing this section, the Technical Strand 3 expert paper (see Carr-Hill, 2015) and the presentations from the subsequent workshop in October 2014 (see OECD, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f) have been taken as the most complete description of the point reached in the planning for how out-of-school children will be accommodated in PISA-D.

The Technical Strand 3 expert paper and the presentations from the October workshop address issues of:

- counting and locating out-of-school children

- finding and identifying out-of-school children
- sampling out-of-school children, persuading them to participate and administering the test
- the design and development of appropriate instruments.

This section uses the same structure, but several subtopics have been merged.

### ***Counting, locating, finding, identifying and sampling out-of-school children***

As mentioned above, PISA makes no attempt to include out-of-school children, and PISA-D is adopting an exploratory approach to this aspect of the project.

Of the reviewed surveys, only PIAAC, STEP, LAMP, ASER and Uwezo include out-of-school children. They achieve this by having target population definitions that are age-based and make no reference to the enrolment/schooling status of individuals.

All five of these assessments sample households. STEP samples households in urban areas only. LAMP, PIAAC and Uwezo sample households across the participating countries, in both urban and rural areas. ASER samples households in rural districts only. In ASER, households are sampled on the day of testing by the test administrators.

Of note from the reviewed surveys is information about how often problems occur with outdated sampling frames, and how these problems are dealt with. These frames may be at the household level or be of sampling units above household level, such as villages, as in the case of ASER.

Note that each of the reviewed household surveys has a reasonably broad age range in the population definition. The narrowest range is Uwezo, with an age definition of 6 to 16 years old. This means that most sampled households will include a respondent within the target age range.

Since the reviewed household-based surveys all sample households and have broad age ranges in their target population definitions, none of them really faces the methodological issues that will be faced by PISA-D as it tries to include children who are both out-of-school and 15 years old. One presentation from the workshop on out-of-school children suggests that PISA-D should adopt direct and indirect aspects. For the direct aspect, an attempt will be made to develop household lists from which 15-year-old out-of-school children can be sampled. For the indirect aspect, attempts will be made to access 15-year-old out-of-school children by contacting employers (OECD, 2014f).

The presentations from the workshop on out-of-school children note that in urban slum areas it may be difficult to establish a household list because it may be difficult to distinguish between households. In this regard, Uwezo tests children in urban areas, some of which would qualify as informal settlements or slums.

Regarding an approach that involves sampling households and testing children in households, as mentioned in the Strand 3 Technical Paper, ASER's procedures are relevant. ASER has specific instructions for test administrators about how to deal with a physical house that seems to house more than one family (as indicated by multiple kitchens). ASER also includes instructions on how to access children who may be shy because they cannot read, and even how to access older children who may not be considered children within their families (see ASER Centre, 2014: 17, 19). This information may be useful for PISA-D.

It may be possible for PISA-D to develop an efficient and effective way to access 15-year-old out-of-school children in households. However, if the survey administration is limited to households (according to the traditional notion), then the project's stated aim of inclusiveness may not be adequately satisfied. This is because any household-based sampling methodology will not reach the most vulnerable and marginalised out-of-school children, because these children are typically not found in households. This issue is addressed in the Strand 3 Technical Paper for PISA-D, which discusses the subpopulations that are omitted from household surveys by design and in practice. Subpopulations omitted by design include: institutionalised populations and displaced populations living in, for example, refugee camps; the homeless; and mobile, nomadic or pastoralist populations. Subpopulations that are under-represented in practice include individuals in fragile, disjointed or multiple occupancy households; urban slum populations; populations living in areas that pose a security risk to visit; and individuals who are marginalised in their households because of illegality or stigma (Carr-Hill, 2015). If PISA-D wishes to collect information about any of these out-of-school children who are typically not found in households, then an alternative method to a traditional household sampling method will need to be employed.

Surveys such as the UNICEF-UIS Out-of-School Children Initiative may be helpful in suggesting approaches to locating, finding, identifying and sampling out-of-school 15-year-olds for PISA-D (OECD, 2014c).

### ***Persuading out-of-school children to participate***

The Strand 3 Technical Paper emphasises the fact that obtaining a sample is only one of the steps that PISA-D will need to consider if it is going to attempt to include out-of-school children in the survey. After out-of-school children have been found, identified and sampled, they must be persuaded to take part in the assessment, and the assessment must be appropriately targeted and appropriately administered.

This fact is also highlighted by the OECD's depiction of the out-of-school aspect of PISA-D as exploratory not only in terms of sampling, but also in terms test and questionnaire administration.

In each of PIAAC, STEP, LAMP, ASER and Uwezo, the survey is administered one-on-one to individuals in each sampled household. None of these surveys offers participation incentives; incentives may be required for PISA-D (OECD, 2014c). One way ASER and Uwezo keep administrative costs down is by using volunteer test administrators. It would most likely not be appropriate to offer incentives to respondents and to use volunteer test administrators.

Survey administration needs to occur at the time at which children are most likely to be available. ASER and Uwezo administer their surveys on the weekends, and PIAAC, STEP and LAMP administer their surveys in the evenings.

It is necessary that administrators obtain the buy-in of parents and children to the survey. At the beginning of an EGRA/EGMA administration, the test administrator reads the child some information about the survey and requests consent. The administration does not go ahead if the child does not give verbal consent (RTI International and International Rescue Committee, 2011). As noted previously, ASER and Uwezo use local volunteers who are more familiar with the area and are better at gaining the trust of the respondents and their families (ASER Centre, 2014; Uwezo, 2014). Uwezo and ASER both impress on their volunteers the importance of approaching the households politely

and giving the household members adequate information about their work before requesting to test children (ASER Centre, 2014; Uwezo Kenya, 2013). Additionally, Uwezo volunteers report children’s results back to parents – in a sensitive way – immediately after the assessment (Uwezo Kenya, 2013).

As discussed in earlier sections, the language of the test administration is significant. In ASER, out-of-school children are allowed to choose the language in which to complete the reading assessment. In Uwezo, all children are allowed to receive the instructions for the mathematics test in whichever language they are most comfortable using.

### ***Administering appropriate test and questionnaire instruments***

It will be essential for PISA-D to use test and questionnaire instruments that are appropriate for out-of-school children. In the context of using appropriate cognitive instruments, a number of points from the reviewed surveys are relevant. EGRA, EGMA, ASER, Uwezo run tests orally and one-on-one, which means that the testee cannot skip tasks. PIAAC, STEP and LAMP all use adaptive testing. In PIAAC, the computer-based assessment is adaptive. In STEP, the second element of the reading test acts as a screening test, and the assessment terminates if the respondent does not pass the screening test. LAMP uses a filter test, which diverts respondents with lower performance and respondents with higher performance to different modules of the main test. ASER and Uwezo also use adaptive testing, but on a much simpler level. In these assessments the administrations starts at a task of middle difficulty, then progresses either up or down depending on how the child performs on that first task. PIAAC, STEP, LAMP, ASER, Uwezo, EGRA and EGMA all include at least some items to test foundational literacy skills and, if they form part of the assessment, numeracy skills as well.

One of the presentations from the workshop on out-of-school children describes how tests used for the out-of-school 15-year-old population need to be targeted at both children who have completed primary school and at children who have never been to school. SACMEQ and PASEC items may be appropriate for children who completed primary school. Items oriented to testing more foundational skills (such as items from ASER or Uwezo) may be appropriate for children who have never been to school. Even more basic pictorial-type items should also be considered for children who have never been to school. A set of channels or gateways will need to be applied to reduce the risk of over-burdening the volunteer test administrators (OECD, 2014e).

In regard to using appropriate contextual instruments, it is relevant that many of the reviewed surveys collect contextual data via interview and observation rather than by leaving the respondent to complete a questionnaire independently. This may be an option worth considering for PISA-D, not only because it may reduce the incidence of missing data, but also because it may be particularly appropriate if questionnaire respondents have limited literacy. In this regard, ASER and Uwezo include questions that address children’s out-of-school status. Also of note is Jangandoo, a survey that was not included in the reviewed assessments but that may be of interest to PISA-D. Jangandoo is a household-based assessment in the ASER model administered in Senegal. Jangandoo includes questions about out-of-school children (asked to their parents).<sup>3</sup>

### ***Implications***

We suggest that the OECD seeks input from ASER and Uwezo, and perhaps the other household-based assessments, to discuss how often they encounter problems with outdated sampling frames and how these problems are dealt with.

We also suggest that the OECD seeks input from ASER (and perhaps Uwezo) about how to deal with multiple-occupancy households, and how to approach children who might be shy because they cannot read, and children who are perhaps considered adults in their households.

In regard to persuading out-of-school children to participate, it would be instructive to review the ways ASER and Uwezo obtain local buy-in to the survey and to consider whether any of their approaches may be applicable for the PISA-D out-of-school children strand.

In regard to administering appropriate test and questionnaire instruments, it would be appropriate for the OECD to investigate an adaptive design for testing out-of-school children; but adaptive test administration can place more demands on test administrators.

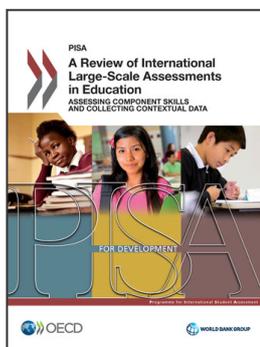
## Notes

1. See [www.tangerinecentral.org/home](http://www.tangerinecentral.org/home).
2. For ASER, many of these partnerships are with district institutes for education and training (ASER Centre, 2014). For Uwezo, the partnerships are with all kinds of institutions that have an interest in education, a belief in citizen agency and a presence in the districts in which they will be administering the survey (S. Ruto, personal communication, 31 August 2014).
3. The Jangandoo household questionnaire can be downloaded from [www.lartes-ifan.gouv.sn/lartes/pdf/ Questionnaire\\_Mu00E9nage\\_enfants\\_MARS\\_2014\\_REMPLI.pdf](http://www.lartes-<u>ifan.gouv.sn/lartes/pdf/</u> Questionnaire_Mu00E9nage_enfants_MARS_2014_REMPLI.pdf). Questions about out-of-school children are on page 8.

## References

- ASER Centre (2014), *Annual Status of Education Report (Rural) 2013*, ASER Centre, New Delhi.
- 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, <http://dx.doi.org/10.1787/5js0bsln9mg2-en>.
- Hungi, N. (2011a), *Accounting for Variations in the Quality of Primary School Education*, SACMEQ, Paris, [www.sacmeq.org/?q=publications](http://www.sacmeq.org/?q=publications).
- Johansone, I. (2012), “Operations and quality assurance”, in M.O. Martin and I.V.S. Mullis (eds.), *Methods and procedures in TIMSS and PIRLS 2011*, TIMSS and PIRLS International Study Center, Chestnut Hill, MA.
- LLECE (2010), *Compendio de los Manuales del SERCE (SERCE Manuals Compendium)*, LLECE and OREALC/UNESCO Santiago, Santiago.
- Martin, M.O. and I.V.S. Mullis (eds.) (2012), *Methods and Procedures in TIMSS and PIRLS 2011*, TIMSS and PIRLS International Study Center, Chestnut Hill, MA.
- Musonda, B. and A. Kaba (2011), *The SACMEQ III Project in Zambia: A Study of the Conditions of Schooling and the Quality of Education*, IIEP, UNESCO, Paris.
- Nkamba, M. and J. Kanyika (1998), *The Quality of Primary Education: Some Policy Suggestions Based on a Survey of Schools: Zambia: An Interim Report*, IIEP, UNESCO, Paris.
- OECD (2014a), Call for Tender 100000990 - PISA for Development Strand A and Strand B, OECD, Paris, <http://tinyurl.com/prjry24>.
- OECD (2014b), “First meeting of the International Advisory Group (IAG) for the PISA for Development project”, summary record, OECD, Paris.
- OECD (2014c), “Approach to sampling and surveying”, presentation at the PISA for Development Technical Workshop on Out-of-School 15-year-olds, Montreal, Canada, 1-2 October 2014.
- OECD (2014d), “Contextual questionnaires”, presentation at the PISA for Development Technical Workshop on Out-of-School 15-year-olds, Montreal, Canada, 1-2 October 2014.
- OECD (2014e), “Developing an assessment framework”, presentation at the PISA for Development Technical Workshop on Out-of-School 15-year-olds, Montreal, Canada, 1-2 October 2014.
- OECD (2014f), “Finding and identifying 15-year-olds”, presentation at the PISA for Development Technical Workshop on Out-of-School 15-year-olds, Montreal, Canada, 1-2 October 2014.
- OECD (2013a), *The Survey of Adult Skills: Reader's companion*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264204027-en>.

- OECD (n.d.), “PISA 2015 technical standards”, [www.oecd.org/callsfortenders/Annex H-Technical-Standards.pdf](http://www.oecd.org/callsfortenders/Annex_H-Technical-Standards.pdf) (accessed 5 August 2015).
- Pierre, G. et al. (2014), *STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills*, working paper, World Bank Human Development Network, Washington DC.
- Results for Development (2015), *Bringing Learning to Light: The Role of Citizen-Led Assessments in Shifting the Education Agenda*, Results for Development Institute, Washington DC, <http://tinyurl.com/p9oe8wp>.
- RTI International and International Rescue Committee (2011), *Guidance Notes for Planning and Implementing EGRA*, RTI International, North Carolina.
- SACMEQ (2015a), “SACMEQ training workshops”, [www.sacmeq.org/training-workshops](http://www.sacmeq.org/training-workshops) (accessed 4 August 2015).
- SACMEQ (2015b), “SACMEQ training modules”, [www.sacmeq.org/training-modules](http://www.sacmeq.org/training-modules) (accessed 4 August 2015).
- SACMEQ (2013), “SACMEQ”, [www.sacmeq.org](http://www.sacmeq.org) (accessed 31 October 2013).
- SACMEQ (2007a), *SACMEQ III: Main Study: Manual for Data Collectors*, SACMEQ, Paris.
- SACMEQ (2007b), *SACMEQ III: Manual for National Research Co-ordinators: Main Study*, SACMEQ, Paris.
- UIS (2009a), *WEI Survey of Primary Schools: Technical Report*, UNESCO Institute for Statistics, Montreal.
- UIS (2009b), *The Next Generation of Literacy Statistics: Implementing the Literacy Assessment and Monitoring Programme (LAMP)*, UNESCO Institute for Statistics, Montreal.
- UIS (2008), “Evaluation of the Literacy Assessment and Monitoring Programme (LAMP) / UNESCO Institute for Statistics (UIS)”, UNESCO Institute for Statistics, Montreal, [www.unesco.org/new/en/unesco/about-us/how-we-work/accountability/internal-oversight-service/evaluation/evaluation-reports/single-view/news/evaluation\\_of\\_literacy\\_assessment\\_and\\_monitoring\\_programme\\_lamp\\_unesco\\_institute\\_for\\_statistics\\_uis/#.VhvCabRVhBd](http://www.unesco.org/new/en/unesco/about-us/how-we-work/accountability/internal-oversight-service/evaluation/evaluation-reports/single-view/news/evaluation_of_literacy_assessment_and_monitoring_programme_lamp_unesco_institute_for_statistics_uis/#.VhvCabRVhBd).
- UIS (2004), “Literacy Assessment and Monitoring Programme (LAMP): International planning report” (draft), UNESCO Institute for Statistics, Montreal.
- Uwezo (2014), *Are Our Children Learning? Literacy and Numeracy across East Africa 2013*, Uwezo and Hivos/Twaweza, Nairobi.
- Uwezo (2013a), *Uwezo Data Cleaning Protocol*, Uwezo, Nairobi, [www.uwezo.net/wp-content/uploads/2012/08/Data-cleaning-and-usage-NOTES1.pdf](http://www.uwezo.net/wp-content/uploads/2012/08/Data-cleaning-and-usage-NOTES1.pdf).
- Uwezo (2013b), “2012/2013 Preparation and training manual – Kenya”, [www.uwezo.net/assessment/training](http://www.uwezo.net/assessment/training) (accessed 10 March 2014).
- Uwezo (2012), *Standards Manual*, Uwezo, Nairobi.
- Uwezo Kenya (2013), “Volunteer workbook - Kenya”, [www.uwezo.net/assessment/training](http://www.uwezo.net/assessment/training) (accessed 11 March 2014).
- World Bank (2013), *National Survey Design Planning Report: Skills Toward Employment and Productivity (STEP): Georgia*, World Bank, Washington DC.



**From:**

## **A Review of International Large-Scale Assessments in Education**

**Assessing Component Skills and Collecting Contextual Data**

**Access the complete publication at:**

<https://doi.org/10.1787/9789264248373-en>

### **Please cite this chapter as:**

Cresswell, John, Ursula Schwantner and Charlotte Waters (2015), "Implementation procedures and approaches to including out-of-school children in educational assessments", in *A Review of International Large-Scale Assessments in Education: Assessing Component Skills and Collecting Contextual Data*, The World Bank, Washington, D.C./OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264248373-8-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to [rights@oecd.org](mailto:rights@oecd.org). Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at [info@copyright.com](mailto:info@copyright.com) or the Centre français d'exploitation du droit de copie (CFC) at [contact@cfcopies.com](mailto:contact@cfcopies.com).