Chapter 4

Contextual data collection instruments used in educational assessments

This chapter looks at the frameworks and instruments for collecting contextual data used by PISA and other large-scale assessments. In the case of each of the reviewed assessments, the chapter outlines the approach used for the following: types of contextual data collection instruments used; mode of delivery; development of contextual data collection instruments; translation, adaptation, verification; main factors and variables used; technical aspects of contextual data collection instruments, such as question formats and scaling and computing of relevant contextual constructs. In each of these areas the implications and lessons for PISA for Development (PISA-D) are identified and discussed.

Chapter 3 of this report reviewed student assessments used by PISA and other programmes. This chapter will review contextual surveys.

One of the main objectives of PISA is to gain data about individual, pedagogical, institutional and systemic factors to describe and compare the contexts of learning, and to investigate the relationships between these contexts and student performance. PISA offers countries the opportunity to collect contextual information from parents (from 2006) and teachers (starting in 2015). Together with the student and school questionnaires, the parent and teacher questionnaires are part of the core instruments for PISA-D (OECD, 2014a: 23).

The purpose of this chapter is to review contextual data collection instruments, at the level of student, parent, teacher and school, implemented by other international and regional surveys – with a view to observing implications for developing countries and in particular for the PISA-D contextual questionnaires.[1] It will also consider the expert paper on context questionnaires by Willms and Tramonte (2014).

This chapter includes the following sections:

- types of contextual data collection instruments used

- mode of delivery

- development of contextual data collection instruments

- translation, adaptation, verification

- main factors and variables, with focus on the seven topics identified as priorities by the participating countries and development partners

- technical aspects of contextual data collection instruments, such as question formats and scaling and computing of relevant contextual constructs.

Types of contextual data collection instruments and mode of delivery

Table D.1 in Annex D gives an overview of the types of contextual data collection instruments used in the international surveys reviewed and their mode of delivery.

PISA uses questionnaires to collect contextual data at the student and school levels. Since PISA 2006, countries can opt to implement a parent questionnaire, and in 2015 an optional teacher questionnaire will be made available to countries. PISA-D intends to implement context questionnaires for students, principals, parents and teachers as core instruments (OECD, 2014a: 23). The mode of delivery envisaged for PISA-D is paper-and-pencil (OECD, 2014a: 37).

The type of contextual data collection tool is largely informed by the survey category (international large-scale, school-based and household-based), which is mainly related to the setting used for the cognitive assessment: group or one-on-one (see Table D.1).

All surveys reviewed collect contextual data. International large-scale surveys use questionnaires for students, teachers and principals. Data from parents are also collected in PIRLS, TIMSS (in 2011) and LLECE. WEI-SPS, which collects contextual data only, uses questionnaires for teachers, principals and curriculum experts. A curriculum questionnaire is also implemented in PIRLS, TIMSS and PASEC.

The school-based surveys EGRA and EGMA, as well as all household-based surveys, are administered in one-on-one settings, allowing the use of interviews for contextual data

collection. EGRA and EGMA provide optional interviews with students, teachers and principals, as well as classroom observation. Household-based surveys focus on individuals in the household, mainly the participant; except ASER and Uwezo where the head of the household is interviewed. ASER and Uwezo combine interviews with observations made in the school or home environment, collecting information from the local government primary school (interview with head teacher) and the village (ASER uses observation only, while in Uwezo the observation is combined with an interview of the local council chairperson or village chief).

Most of the questionnaires and interviews used for contextual data collection in the surveys reviewed are administered in paper-and-pencil mode, while delivery of questionnaires in PISA will be largely online from 2015 onwards (except for PISA-D and countries using the paper-and-pencil assessment option). Of the other assessments reviewed, only PIRLS and TIMSS offer an online questionnaire option for teachers and parents. PIAAC and STEP are the only household-based surveys that use computer-assisted interviews.

Implications

In regard to the questionnaire type, Willms and Tramonte (2014: 20) underline the importance of discerning the best informant for measuring the relevant constructs. The authors argue that implementing a parent questionnaire would be a useful option to collect data on family issues for PISA-D. The comparison of international surveys shows that parent questionnaires are mainly used in large-scale international surveys with younger student populations (Grade 4 in PIRLS and TIMSS; Grades 3 and 6 in LLECE) as well as in the household-based surveys ASER and Uwezo, where the head of the household is interviewed in a one-on-one setting. In this regard Willms and Tramonte (2014: 20) suggest to consider an interview approach for parents in PISA-D, which would be valuable to assess parent's literacy skills and employment, similar to the approach of household-based surveys with an international focus (LAMP, STEP, PIAAC).

While Willms and Tramonte have highlighted the importance of discerning the best informant, a major consideration is the cost-benefit ratio of parent questionnaires, given the effort needed to carry them out. This is especially relevant of an interview approach, as securing response rates through one-on-one interviews is a financial burden. This must be weighed against the benefit of such data. Comparisons between student and parent questionnaire responses to family-related questions in PISA have shown that students are a reliable source of information for family-related questions such as parents' occupation, occupational status, language, parental education and so on.

A teacher questionnaire is carried out in all large-scale international surveys as well as in most EGRA and EGMA administrations, regardless of whether students are sampled from intact classes in schools (PIRLS, TIMSS, LLECE, EGRA, PASEC) or randomly within schools (PISA, SACMEQ, PASEC, EGMA). A teacher questionnaire is used throughout international surveys to assess the following key areas: quality of instruction, school resources, language at home and in school, and learning time.

Willms and Tramonte (2014: 20) support the use of a teacher questionnaire if many of the classroom and school constructs could be better addressed by teachers than by students or principals. For developing countries, a teacher questionnaire has potential benefits, compared to collecting the more aggregated school-level data through the principal questionnaire. For PISA-D, it is worth remembering that the student sample in PISA is not class-based: PISA is seen as an accumulation of the student's educational

experience. Drawing conclusions about teacher background and strategies is more difficult for PISA than for a class-based assessment.

Regarding the mode of delivery, electronic means such as tablets are worth considering, as noted in the discussion of test design in Chapter 3 of this report. This option would allow spoken and visual language components to be incorporated for struggling readers. Electronic delivery offers a potentially wider range of options for collecting contextual data, as well as for handling and processing data.

Development of contextual data collection instruments

Table D.2 in Annex D gives an overview of the main bodies involved and the main steps in the process of developing the different contextual data collection instruments, including review options and piloting/field trialling. Translation, adaptation and verification processes, also key elements of the development process, are described separately.

Theoretical conception of contextual data collection instruments

Questionnaire development in PISA is based on a context framework. This outlines the theoretical and scientific background of the questionnaire content to be measured, and of the interactions and relationships between certain factors and student achievement, as well as important non-cognitive learning outcomes. The PISA context framework (OECD, 2013a, n.d.-a) is based on two approaches: *i)* a model of learning by Carroll (1963); and *ii)* a policy framework that addresses questions of relevance to participating countries (Willms and Tramonte, 2014: 4).

The factors defined in the framework are structured in a two-dimensional taxonomy of educational outcomes and predictive factors (OECD, 2013a: 175). This taxonomy is based on research in educational effectiveness of input, process and outcome measures at the system, school, classroom and student levels. The basic structure of this taxonomy is derived from the "input-process-outcome model" that was developed in the 1960s for the IEA (Purves, 1987). In PISA this model has been expanded with the different levels on which contextual factors affect student learning (system level, school level, classroom level and student level).

The factors can further be classified as domain-independent or domain-related measures. The domain-independent measures include (Willms and Tramonte, 2014: 3, 4):

- student-level inputs, such as grade, gender, parent occupation and education and migration background

- classroom instructional processes, such as learning time, disciplinary climate and teacher support

- school-level contexts, such as school type, school size, class size, school resources and learning environment, human resources, school location and community size

- school-level processes, such as school climate, teaching practices, assessment and evaluation policies, and professional development

- non-cognitive outcomes, such as truancy, engagement and sense of belonging.

Domain-related measures include, for example, attitudes towards mathematics, reading or science, motivation and self-concept in mathematics, reading or science, and instructional practices in these subjects, some of which are classified as processes (such as instructional practices) and some as outputs (such as attitudes).

PISA-D aims to extend existing constructs and scales derived from these factors in a way that makes them more relevant for contexts found in economically developing countries. Section 4.3 of this report discusses the main factors relevant for this review.

Theoretical conception of contextual data collection instruments varies across the other assessments reviewed. Most of the international surveys reviewed state a theoretical underpinning of the context factors collected, as well as their relationship of these factors with achievement. This theoretical underpinning combines both educational research questions based on a model of learning and policy questions.

Large-scale international surveys

The context frameworks for PISA (OECD, 2013a), WEI-SPS (Zhang, Postlethwaite and Grisay, 2008) and PIRLS and TIMSS (Hooper, Mullis and Martin, 2013) are highly elaborated. As mentioned above, the primary theoretical conception in OECD and IEA-led studies is based on the input-process-outcome model (Purves, 1987), where input, process and outcome factors are located on student, classroom, school, and system level – representing a two-dimensional taxonomy of educational outcomes and predictive factors.

More specifically, PIRLS and TIMSS classify context factors for national and community contexts, school contexts, classroom contexts, and student characteristics and attitudes. PrePIRLS and TIMSS-Numeracy are consistent with the PIRLS and TIMSS frameworks (and use the same context questionnaires).

In WEI-SPS, the principal indicators are organised into: contexts (the environments in which individual schools operate); inputs (material and human resources available to schools); and processes (indicators/processes outlined at both the school and classroom levels) (UIS, 2009a). The context frameworks for PISA, PIRLS and TIMSS and WEI-SPS (embedded in the UIS (2009) technical report) are publicly available.

The large-scale international surveys SAQMEC, PASEC and LLECE use analytical models to describe the context factors collected and the expected relationships with achievement. Similarly the analytical models consider different levels: student and family, classroom, school, and system/national/community level. These models are usually described in technical or results reports or technical documentations (CONFEMEN, 2012; Dolata, 2005; LLECE, 2009).

In SACMEQ, the context questionnaires use a general two-level model, which is based on existing literature on student learning, especially Carroll's model of school learning (Carroll, 1963) and Creemers' model of effective classrooms (Creemers, 1994; Hungi, 2011a: 5). The model was hypothesised for factors influencing student achievement in reading and mathematics, with students located on level one and schools on level two (students within schools) (Hungi, 2011a). Three categories of variables were hypothesised to directly influence achievement at the student level: individual characteristics, personalised learning support and home environment. Four categories of variables were hypothesised to directly influence achievement at the school level: teacher characteristics, classroom environment, school head characteristics and school environment (Hungi, 2011a, 2011b).

PASEC reports educational indicators at three levels: the socio-economic background of students, teaching conditions and policy guidelines. These indicators are matched with students' competencies (CONFEMEN, 2013). The analysis scheme involves individual and familial student background characteristics as well as early learning opportunities as antecedents (see, for example, the national report for Chad (CONFEMEN, 2012: 88). In addition to these, the following factors are expected to affect student achievement: personal schooling conditions (for example, owning school books), profile of the school principal, school characteristics (such as electricity, rural or urban location), profile of the teacher (such as qualifications, years of teaching experience, gender), class characteristics (such as class size) and pedagogical organisation (such as multi-grade, double shifts).

Questionnaire development in LLECE broadly emphasises factors associated with student achievement that can be directly affected by education systems. The macro conceptualisation is guided by five strategic aims of the Regional Education Project for Latin America and the Caribbean (PRELAC), an association of education ministries, in order to support progress towards Education for All (LLECE, 2009: 35). Three principal theoretical domains are covered by the questionnaires: socio-cultural characteristics, educational opportunities, and academic achievement. The questionnaires also cover the transversal domain of educational equity (LLECE, 2009). SERCE questionnaire development also considered findings from SACMEQ, PISA and TIMSS for important context factors that are expected to affect achievement.

School-based surveys

Contextual data collection instruments in EGRA and EGMA are developed by RTI and are based on the Snapshot of School Management Effectiveness instrument (SSME) (Crouch, 2009). The SSME instrument is in turn based on reviewed literature (Lockheed and Verspoor, Henneveld, Schiefelbein and Wolff, Moura Castro – Crouch, 2008). SSME comprises five major domains: *i)* pedagogical leadership and management; *ii)* class and classroom management; *iii)* school management; *iv)* parent and community involvement in the school; and *v)* district and system-level support and supervision. The factors are located at student/family, teacher/classroom, school and community level. Implementation of the SSME contextual instruments is optional for each EGRA/EGMA implementation, and in some instances countries do no implement any questionnaires from SSME at all.

Household-based surveys

The household-based adult literacy and skills studies PIAAC, STEP and LAMP focus on factors related to adult literacy as well as work-related skill acquisition and use.

Contextual data collection in PIAAC is based on three main policy questions that are further theoretically underpinned: *i)* how skills are distributed; *ii)* why skills are important; and *iii)* what factors are related to skill acquisition and decline. PIAAC collects a range of information on the factors which influence the development and maintenance of key skills, such as education and training, current status and work history, current work and last job (for those currently employed or self-employed, who have worked in the last five years), social background, language, engagement with literacy, numeracy and information and communication technologies (OECD, n.d.-b) (Allen et al., 2013: 42). Additionally the background questionnaire includes a module called the "job requirements approach". This module collects a range of information on the reading and numeracy-related activities and technology use of respondents at work and in everyday

life, and on the generic skills required of individuals in their work. Respondents are also asked whether their skills and qualifications match their work requirements and whether they have autonomy over key aspects of their work (OECD, 2013b: 40).

STEP aims to provide data about skill stocks and job demands in low- and middle-income country contexts and focuses on work-related skill acquisition, use, and distribution. STEP uses "a multidimensional concept of skills that goes beyond educational attainment to capture human capital more comprehensively" (Pierre et al., 2014: 7). The STEP survey consists of a household survey and an employer survey. Both contain detailed measures of required education and experience and of the required skills in reading, writing, math, problem-solving, interpersonal and socio-emotional traits, technology use, and manual work required by jobs (Pierre et al., 2014: 2, 9). The household survey contains seven modules of contextual data collection instruments. Module 1 comprises a household roster and dwelling characteristics, and aims at getting a full picture of the household and its members that could influence the outcome of interest (such as obtaining a job) for the individual who will later respond to the full questionnaire. The section about dwelling characteristics includes household assets, from which an asset index is constructed to be used as a proxy for wealth (Pierre et al., 2014: 14). Modules 2 to 7 are part of the individual questionnaire and collect data on education and training, health, employment, self-reported cognitive skills and job-relevant skills, personality, behaviour and preferences, and language and family background. A detailed description of the questions module-by-module is available in Pierre et al. (2014).

Contextual data collection in LAMP focuses on more general factors influencing adult literacy skills. To direct the contextual data collection, research questions cover five major areas that are of interest to policymakers: population distribution of literacy skills, antecedents of literacy skills, relationship of literacy skills to social environment, relationship of literacy to other proxy variables and monitoring trends in literacy skills (UIS, 2006). The background factors are structured into classification variables (used to identify subpopulations), relationship variables (such as expected relationship with literary levels, skills acquisition, enhancement and maintenance) and profiling variables (to statistically profile groups with particular levels of skills acquisition) (UIS, 2006).

For ASER, it is reported that household indicators are recorded "in order to link education status of the child with the household's economic conditions" (ASER Centre, 2014: 21). Moreover, ASER implements the Right to Education indicators; the "4A-framework" of Right to Education is closely linked to international human rights law and covers availability, accessibility, acceptability and adaptability (ASER Centre, n.d.). In any discussion about ASER tools, it is important to keep the basic objective of the exercise in mind. ASER is primarily an attempt by citizens to understand the status of schooling and basic learning of the children in their district. The tools are aligned to achieving this objective. The biggest challenge in ASER is to make the tool as simple as possible without sacrificing rigour (Banerji and Bobde, 2013).

For Uwezo, no documentation is publicly available of the theoretical background for contextual data collection and expected relationships with achievement.

Development process and main bodies involved

Development process

Most large-scale international surveys follow a very similar questionnaire development process as PISA. In most cases policy priorities and/or research questions

are defined, then further outlined in a "context framework" that provides theoretical underpinning of the context variables and factors implemented in the survey, as well as their relation to achievement. This is the process used by PISA, PIRLS, TIMSS, WEI-SPS and PIAAC. Alternatively, SAQMEC and LLECE construct analytical models to describe the relationships of contextual factors surveyed and achievement. PASEC's models are not explanatory but rather descriptive.

New items are developed based on these priorities and research questions. Generally throughout the surveys, the contextual data collection instruments are updated from one cycle to the next, to include topics of high (policy) relevance. In large-scale international surveys such as PISA, TIMSS and PIRLS, this is a balancing act between maintaining consistency with former measurements to report on trends and considering recent developments and current policy priorities. Sometimes where a large amount of new and additional material would considerably increase the response burden for participants, existing questionnaire items are retired to make way for the new; thus, not increasing the total amount of material. This has happened in PIRLS and TIMSS (Mullis et al., 2012a: 3).

The development of contextual questionnaires follows a typical sequence. First, developers revise existing material, such as frameworks, analytical conceptions and items. New material is created, through consortia, expert groups and policy input. This is followed by a review phase, through governing authorities, donors, participating countries and national experts. Revisions are made, through consortia and expert groups. Field trialling and data analysis take place, with ensuing revisions and reviews. All this input is reflected in the final decisions about questionnaire design and item selection for the main survey. Review activities ensure appropriate coverage of the topics specified in the contextual frameworks and analytical conceptions, the analytic potential of the items and reporting scales, the clarity of the items, and suitability of the items in the respective national context. These processes are similar for all international large-scale surveys as well as the household-based surveys PIAAC, STEP and LAMP.

A key element of the contextual data collection instrument development process is piloting or field trialling of (new) items, questionnaire designs and administration procedures. A pilot is considered as a pre-study to a field trial, with a smaller sample and often with a focus on specific research questions – as opposed to a field trial, which usually is to test the whole assessment, including instruments and item functioning, as well as procedures.

Piloting or field trialling of contextual instruments is a main part of all large-scale international surveys as well as the household-based surveys PIAAC, STEP and LAMP.

In EGRA and EGMA two pilots were conducted to validate the Snapshot of School Management Effectiveness instrument developed by RTI (Crouch, 2009). The report for an EGRA and EGMA implementation from Morocco in 2011 states:

> *Each instrument was pretested in eight schools within the region of Doukkala Abda. (These schools were not included in the sample used for final assessment.) The SSME instrument was then reviewed in light of the pretesting experience, any phrasing of questions that led to misunderstandings was clarified, and problematic questions were removed or modified. (Messaoud-Galusi et al., 2012: 27)*

For ASER, a small pilot during test administrator training is conducted, but no information is available about revisions of contextual data collection instruments resulting from this pilot.

For Uwezo, the regional office undertakes tool development and then the national offices review the tools to ensure the items' relevance. For example, the annual plan and budget document for 2013 for Uganda refers to a survey tools review meeting during which country-specific adaptations to the tools were to be reviewed and adopted (Uwezo, 2013: 28). The Uwezo survey tools are involved in piloting activities in all three countries. An inspection of the pilot report for 2013 for Uganda shows, for example, questions that were frequently incorrectly coded by test administrators; such as where administrators ticked an open question instead of writing a number.

Main bodies involved

The review of international surveys found that for all of them, various bodies are involved in questionnaire development (see Table D.2 in Annex D). The extent to which the different bodies have an influence on the development – and during which phases (especially during the theoretical and/or analytical conception) – could not be determined during the review of international surveys.

Involvement at a policy level similar to the PGB, or stakeholder level, is explicitly reported for PIAAC (Board of Participating Countries), SACMEQ (country ministers of education), PASEC (CONFEMEN), LLECE (supporting the aims of PRELAC, an association of education ministries in Latin American and the Caribbean) and WEI-SPS (stakeholders, project steering committee). For STEP and LAMP, the World Bank/UNESCO Institute for Statistics (UIS) define overall priorities and participating countries may contribute specific priorities. Contextual instrument development in PIRLS and TIMSS as well as EGRA and EGMA is primarily based on research (learning theories and models). In the case of PIRLS and TIMSS, "policy interests" are integrated through national research co-ordinators as the main reference source for framework development; similar to EGRA/EGMA, where specific interests of RTI and donors such as the United States Agency for International Development (USAID) and the World Bank (in the case of EGRA) may have an influence. For ASER and Uwezo there is little documentation of the bodies involved in contextual data collection instrument development. Broadly the content seems to be based on common assumptions about relationships between specific contextual variables and achievement. Specific interests of the assessment centres and/or donors may play a role in the development of contextual data collection.

There is a questionnaire expert group, similar to PISA's, for PIRLS and TIMSS, SACMEQ, LLECE (for TERCE, labelled a high-level technical advisory board), WEI-SPS (OECD-led), EGRA and EGMA, PIAAC (OECD-led) and STEP.

On the operational level, participating countries are involved generally through the national centres responsible for the implementation of the survey in the country (such as through national project managers in PISA, national research co-ordinators in PIRLS and TIMSS, country co-ordinators in LLECE) and national experts. Similar to PISA, these country representatives review the questionnaires and give feedback on specific content and its fit with the context of the national education system.

Implications

Building on more than 15 years of experience with framework and questionnaire development, the processes and theoretical concepts applied in PISA are highly elaborate. In regard to options for PISA-D, the main consideration is to underline the importance of involving participating countries at the policy level, research level, as well as the operational level. On the policy level, country representatives (including government, non-government and donor representatives) need to be involved in order to ensure that relevant national education policy issues are identified and covered in the data collection. In addition to that, national education experts (researchers, teacher trainers, school principals and teachers) need to be involved in order to identify questions of particular education research interest. The comparison of contextual frameworks shows that policy interests and research interests complement each other and need to be considered when conceptualising background contextual questionnaire development.

Education experts and national project managers also need to be involved in reviewing and adapting the questionnaire constructs to make them meaningful for the country context. This is of particular importance for the seven priorities identified for the context questionnaires (Willms and Tramonte, 2014), and specifically for measures of socio-economic status and school resources (see section 4.3).

With respect to extending existing questionnaire scales and introducing new constructs that are of relevance to developing countries, PISA-D needs to find a good balance between "core" and "new" content, in order to not increase response time and thereby the response burden on participants.

It is also important to provide options and assistance to PISA-D countries to cover topics that are of particular national interest – and that would otherwise not be considered in the international PISA questionnaire – in national questionnaires or other national context data collection instruments.

Field trialling of context questionnaires is a standard procedure for any PISA administration. A field trial will be essential to PISA-D to allow for improvements as necessary to new questionnaire constructs, extended scales and implementation procedures.

Translating, adapting and verifying contextual data collection instruments

In the literature reviewed for the international surveys, the process of translation and adaptation is mainly described for the cognitive instruments and these processes of translation, adaptation and verification apply as well for the contextual data collection instruments. The process is therefore only briefly described in this section.

Table D.3 in Annex D gives an overview about the source version and translated languages for the contextual data collection instruments, and the translation, adaptation and verification procedures applied in the different surveys.

Languages

Usually the language of assessment is used for the contextual data collection instruments. In the surveys reviewed, most of the contextual questionnaires are developed in English. Surveys with the most diverse and the greatest number of languages are PIRLS and TIMSS (58 languages; with 11 languages alone in South Africa), PISA

(46 languages including right-to-left and top-to-bottom script), and PIAAC (about 30 languages). LAMP, STEP and WEI-SPS also cover a broad range of languages.

Specific information was available for WEI-SPS on the language of contextual data collection instruments. In countries where one language is primarily used in the education system, surveys were translated into that language only. In countries where more than one language of instruction is used in primary grades, either throughout a country or even within a school, the questionnaire needed to be delivered in a language in which the school principals and teachers would be proficient (that is, the language used in teacher training) – which does not necessarily match the language in which the students are taught. For multilingual countries, surveys were therefore translated into the national language, *and* into languages in which teachers and principals were expected to be proficient (UIS, 2009).

In PIRLS and TIMSS it is reported that in some countries where the language of instruction differs from the language used at home, countries translate the parents' questionnaire into one or more additional languages (the languages most commonly spoken in the home), to allow parents to fill out the questionnaire in the language they feel most comfortable using (Yu and Ebbs, 2012: 4). In PIRLS/prePIRLS in South Africa, the teacher and the principal questionnaires were administered in Afrikaans or English only (not in all 11 official languages), based on the assumption that most teachers and school principals would have been able to speak, write and understand these languages, as required by their teacher training qualifications (Howie et al., 2012: 24).

Language (language of instruction and language spoken at home) as an important context factor is discussed in more detail in section 4.4.2.

Translating, adapting and verifying

Generally among the surveys a source version is provided for translation. Often this is in English, but not always: French in PASEC; Spanish in LLECE; English and French in PISA; English, French and Spanish in LAMP. In PISA a double independent translation (either from the English and French source version or from the English source version), followed by a third independent reconciliation, is used. This procedure is also used in PASEC and SACMEQ (reconciliation can be carried out by the two translators or a third person), WEI-SPS, PIAAC and STEP. PIRLS and TIMSS use a similar regulation, indicating that the translation is considered correct if more than one translator is used. LLECE uses a three-step translation of Spanish into Portuguese, then the Portuguese version is back-translated into Spanish, and is then compared and validated.

Guidelines for translation usually include rules for adaptations and verification. Standardised procedures are provided in most of the international large-scale surveys as well as the household-based surveys that aim for international comparison (PIAAC, STEP and LAMP). In EGRA/EGMA, ASER and Uwezo the countries are responsible for translating, adapting and verifying local versions, and only little information is available about the procedures.

Most surveys acknowledge the importance of national adaptations for questionnaires to match national contexts, as described for example in LAMP: *"The adaptation of the background questionnaire is of utmost importance as it will provide key elements for analysis and, therefore, for accomplishing the goals set at the national level"* (UIS, 2009b: 37).

Usually any adaptations need to be documented and approved by the responsible authority. Adaptation can be highly complex, as noted for PIRLS:

Adapting the questionnaires to specific educational contexts is quite complex, particularly for countries that administer the survey in multiple languages or at a different grade than the internationally defined target grade ... Verifiers received detailed instructions and information on each country's participation configuration to ensure appropriate review and relevant feedback on the national materials. (Yu and Ebbs, 2012: 2)

This shows how important it is to establish clear guidelines for adaptations and a thorough verification process.

No other specific translation or adaptation issues that would be of particular relevance for contextual data collection instruments in PISA-D were reported or identified during the literature review. Some of the documents reviewed contained general information about country adaptations to questionnaires, which are comparable to those made during the common PISA translation and adaptation process (see, for example, Foy, Arora and Stanco, 2013; UIS, 2009a: 129-153).

With respect to national adaptations, options for including specific national questions play an important role. National questions can help deepen specific areas of interest that would otherwise – in an international context – not be included. Surveys that explicitly encourage countries to include national options in the contextual data collection instruments are PISA, PIRLS, TIMSS and LAMP. The number of questions of national interest to the questionnaires however is limited, to not put an additional burden on the respondents (especially students). EGRA/EGMA, ASER and Uwezo also allow for adding questions of national interest.

Implications

For multilingual countries, it is important to consider which languages are the most appropriate for the different groups of respondents. Questionnaires are preferably translated into languages in which students, teachers, principals and parents are expected to be proficient; these languages do not always match with the defined "language of assessment". Identifying the appropriate language can be a challenge, especially in multilingual countries, where the language of instruction differs from the language most commonly spoken at home. We suggest collaborating closely with national centres of PISA-D countries to carefully assess the language situation for the different groups of respondents and to determine the most appropriate languages for each group. One example is PIRLS and TIMSS, where the parent questionnaires are translated into the languages most commonly spoken in the home.

In relation to translation, standardised guidelines and procedures play a crucial role for ensuring high quality of translations and hence comparability across languages. PISA has established high standards for translation, adaptation and verification and is constantly improving these processes. In PISA 2012, instruments were translated into 46 languages (for 98 national versions) including right-to-left scripts (Arabic) and top-to bottom scripts (Chinese traditional and simplified script). This testifies to the wide range of translation experience that has already been gained in PISA.

In order to best accommodate national contexts, national adaptations are of particular importance to contextual data collection instruments. Procedures that ensure international comparability of adapted content are a key element of the translation process. PISA has

highly developed and well established procedures to achieve this. Key elements of the adaptation and verification process that are typical for large-scale international surveys are:

- highlighting content that requires adaptation in the draft versions of the questionnaires

- accurately documenting adaptations

- verifying that national adaptations are appropriate

- verifying the quality of the translation in regard to mistranslations, undocumented deviations, and linguistic equivalence to the source versions.

In addition to verification procedures, field trial analyses (such as frequency reports and scale functioning within and across a country) help to detect translation and adaptation issues early, and to revise items and scales for the administration in the main survey.

The review of international surveys shows clearly that the translation, adaptation and verification procedures applied in PISA comply with very high standards.

In order to best accommodate national interests, it is important to facilitate national options – allowing countries to include specific questions concerning policy priorities and other topics that are of particular importance to a country, and that would otherwise – in the international context – not be covered.

Main factors and variables for PISA-D

The PISA context framework defines and explains the content to be measured with the PISA context questionnaires (see, for example, OECD, 2013a). In order to make PISA more relevant for contexts found in developing countries, Willms and Tramonte (2014) identified seven key topics in which the PISA context questionnaires should be enhanced for PISA-D. These seven topics are based on the two approaches of the PISA context questionnaire framework (model of learning by Carroll (1963), policy framework; see section 3.1.1) as well as on consultations with the participating PISA-D countries:

1. early learning opportunities – details such as whether students attended an early years learning or care programme or repeated a grade during their early years

2. language at home and school – information on students' familiarity with the language of the test

3. family and community support – measures of parental involvement, social capital and cultural capital; measures that focus directly on the role and involvement of other community members in the school, including in relation to school safety and security, and descriptive variables regarding types of community

4. quality of instruction – extent to which class time is spent in independent activities, such as working in workbooks, versus small group activity and whole-class teacher-centred instruction

5. learning time – students' learning time in and out of school; school attendance and students' participation in the labour market

6. student socio-economic status – extending the current indicators of the PISA index of economic, social and cultural status to include more items at the lower end of the socio-economic scale; 'poverty-related' measures

7. school resources – quality of school environment, school infrastructure, student and staff safety.

These seven key topics have been included in the terms of reference of the PISA-D call for tender (OECD, 2014a) for context questionnaires. The themes were also discussed during the PISA-D technical workshop in Washington in April 2014, and participating countries and partners agreed on suggestions for enhancement and modification (OECD, 2014a: 52).

The following comparison of main factors and variables in international surveys focuses on these seven key areas. This comparison does not discuss other core factors and variables that have been included in PISA and that are not likely to require much modification for PISA-D (such as student demographics including age, gender, and migration background) (Willms and Tramonte, 2014).

Table D.4, Table D.5, Table D.6 and Table D.7 in Annex D give an overview of the factors used in international surveys relevant to PISA-D and discussed in this section. The sources used to identify the main factors for each survey (questionnaire frameworks, technical and other reports, specific context questionnaires) are stated in the respective tables' footnotes. The detailed references are listed in the references at the end of each chapter and annex of the report. All questionnaire material used – as well as specific items referred to – are publicly available, and therefore no example items have been added to the annex.

Early learning opportunities

Students' performance captured in PISA represents the cumulative result of children's learning experiences in school and outside of school, in the family and community. Gathering information about the early learning opportunities of students is important for PISA-D in order to obtain a broad picture how this may have informed teenagers' learning experiences. Wills and Tramonte (2014: 5, 8) argue that children's early learning experiences differ substantially within the partner countries and compared to OECD countries, and are likely to play a more dominant role in economically developing countries than in OECD member countries, especially in rural areas.

Information about students' early learning experiences is collected in PISA through student and parent questionnaires.

PISA collects information about students' attendance during early education, age at the beginning of primary education, and grade repetition during primary and lower and upper secondary education. The latter is also captured in the parent questionnaire.

Most other large-scale international surveys also collect data on early learning. (The exception is WEI-SPS, which only collects data from teachers and schools.) PIRLS and TIMSS collect data on early learning from parents; SACMEQ from students; PASEC from students; and LLECE from students and parents. EGRA and EGMA collect information from students; ASER and Uwezo from heads of households; and STEP from respondents in the target population only.

An interesting supplement for PISA-D would be the PIRLS and TIMSS home questionnaire, the Learning to Read Survey. This survey is directed to parents and

includes information about reading activities before primary school, early literacy and numeracy activities, as well as reading and quantitative readiness at the beginning of primary school (Martin, Mullis and Foy, 2013: 67, 68).

LLECE also asks about early reading, and how often someone at home reads aloud to the child. The respective questions are administered to students in Grades 3 and 6 as well as to parents.

ASER and Uwezo collect information about the pre-school status (for children under five) and school status of the child in the interview with the head of household. Of particular interest to PISA-D may be the inclusion of questions about "out-of-school-status". This refers to children aged 5 to 16 who are currently not enrolled. Data is collected about whether the child was never enrolled or dropped out; and if the latter, that schooling status when the child left the school, and the year they dropped out.

Language at home and school

Language at home and school is an important indicator in regard to equality. In several economically developing countries students are taught in a different language than their first language. Also, in some countries, the language of instruction changes between different levels of education (such as between third and fourth Grade) (Willms and Tramonte, 2014: 8).

All international surveys reviewed include questions about language.

In the PISA student questionnaire, the language spoken at home is one of the core variables. Additional questions about language were included in the educational career questionnaire in PISA 2012, and in the school questionnaire in PISA 2009. These questions related to students in the national modal grade, which is the year level attended by most 15-year-olds in a country. These questions collected data on the proportion of students in the national modal grade that had a first language other than the test language; and options for students in the national modal grade whose first language was not the test language. The main aim of these questions was to gain more detailed information about students whose first language differs from the language of assessment (which for PISA is the language of instruction).

PIRLS and TIMSS ask the students about the frequency of speaking the test language at home. In addition, the parent questionnaire asks about the language most often used at home, and the language spoken by the child before starting school. Of note is that PIRLS and TIMSS ask if the books at home are mainly in the test language. (PISA collects data about students' books at home as an indicator for socio-economic status.)

On classroom level, PIRLS and TIMSS ask the teacher about the number of students that have difficulties in understanding the test language as it is spoken. At the school level, PIRLS and TIMSS collect data about the proportion of students for whom the test language is their native language, and provisions for reading instructions in their mother tongue for students with a different mother tongue to the language of the test.

PIAAC asks about the language most often used at home, and any second language learned.

STEP and LAMP provide a full picture of the languages that dominate in the household. These language questions focus on the languages that respondents speak, read and write to a level that would enable them to use the language in a job. Questions are about mother tongue (the first language the person learned), the language that is mainly

spoken in the house, the total number of people in the household that speak any of the official country languages (speaking does not necessarily include reading or writing), languages in which the respondents speak, and in which they read and write well enough to work in a job that requires that language (Pierre et al., 2014: 33, 34).

Other assessments focus more on language use at school. PASEC asks about the languages spoken by the teacher. Similarly, EGRA and EGMA ask about the native language of the teacher. LLECE collects data on whether the language of instruction is used for partial or all of instruction; and about indigenous language services and resources. WEI-SPS asks about the official time used for teaching the language of instruction. Uwezo includes questions about the number of textbooks provided for language instruction in English and Kiswahili.

Socio-economic status

Measures of SES are included in every survey. Various assessments use a range of factors and variables, frequently including education and occupation, to construct SES- and poverty-related measures. The derived constructs and scales to measure SES and/or poverty are described in more detail in section 4.6.

Socio-economic measures in PISA include factors such as parental education, occupation, employment status, home possessions and home educational resources, including books at home. These factors are captured on the student level. In the parent questionnaire, parents are asked about their education, occupation, the annual household income and parents' educational expectations for their child.

Similar factors are used in PIRLS and TIMSS (focusing on the parent questionnaire for Grade 4 students). On the school level, PIRLS and TIMSS include a question about the SES of the school's immediate area (high, medium or low).

Specific indicators relevant to children living in poverty are included in SACMEQ, PASEC, LLECE, EGRA and EGMA, ASER and Uwezo.

The SACMEQ student-level data collection includes questions about number of siblings, number of meals per week, household tasks, learning culture at home, if the parents are alive, if the child is living with parents or relatives, and about the home environment.

In PASEC, students are asked about their standard of living (poor, intermediate, rich) and if mother and father are literate (able to read and write).

LLECE includes questions about parental education, such as whether the student's mother and father can read and write. Measures of home utilities include availability of electricity, water and sewage services; the construction materials of the home; and the availability of a phone and cable or Internet.

EGRA and EGMA, ASER and Uwezo all use similar variables for home possessions/facilities. Additionally, EGRA and EGMA include questions about the type of toilets, method for cooking, and water source for washing. ASER includes questions about parental education, asking about school attendance and status of completed education, and also asks about the availability of reading material and if anyone in the household knows how to use a computer. Uwezo additionally captures the main source of household income; if lighting is available in the house; the number of meals per day; about possessions such as a radio, TV, computer or mobile phone; about livestock such as

cattle, donkeys, camels, sheep or goats; and about transport such as a bicycle, motorbike or cart.

Indicators for socio-economic status in household-based surveys focusing on the adult population (PIAAC, STEP and LAMP) are mainly based on household characteristics and facilities such as quality of housing, equipment available in the household, a common set of assets, and land and livestock ownership. In STEP, a particular asset index (Pierre, et al., 2014: 15) is used as a proxy for wealth. In LAMP, information about household facility and living environment is used to create measures of SES, which are then classified into four socio-economic groups: affluent or well-off, comfortable, poor, or subsistence level.

PIAAC, STEP and LAMP obtain basic information about employment, such as the labour force status (employed, unemployed, or inactive; including self-employed – with and without pay; underemployed, or holding low-productivity jobs). For those who work, STEP inquires in detail about their occupation, earning, hours worked, and so on. Pierre et al. (2014: 22) indicate that a large proportion of the labour force in developing countries is self-employed, underemployed or holding low-productivity jobs. For the self-employed (with and without paid work), the survey therefore asks a series of specific questions that help determine the overall success of their businesses and to find out the extent to which such work is voluntary (such as by asking about the preference for wage jobs versus self-employment). Questions related to education and training in STEP are aimed at obtaining a full picture of the acquisition of skills throughout the respondent's lifetime. The module does this by asking questions related to formal education, lifelong learning, and other types of training and certificates (Pierre, et al., 2014: 16).

UNESCO's International Standard Classification of Education (ISCED) is used to classify education in PISA, PIRLS and TIMSS, WEI-SPS, PIAAC, LAMP and STEP.

The International Standard Classification of Occupations (ISCO) is used to classify occupation in PISA and PIAAC (ISCO 2008), PIRLS, TIMSS, LAMP and STEP.

Quality of instruction

Quality of instruction is important to PISA-D as it varies widely from OECD and partner countries, especially in rural schools (Willms and Tramonte, 2014: 8). Measurements of classroom context and quality of instruction in PISA are mainly related to time spent on activities in connection with "direct instruction" and cognitive activation (Klieme et al. 2009, in Willms and Tramonte, 2014: 11). Wills and Tramonte (2014: 11) assume that for the context of PISA-D, measuring more basic instructional activities may be useful.

Measures for quality of instruction are included in all large-scale international surveys, in EGRA and EGMA, and in ASER. These measures can be categorised as general aspects and domain-related aspects of quality of instruction.

From the surveys reviewed, a number of general aspects of the quality of education are considered relevant for the context of PISA-D:

- The PASEC teacher questionnaire asks about pedagogical practices. PIRLS and TIMSS teacher questionnaires cover instructions to engage students in learning. WEI-SPS asks teachers about active learning. EGRA and EGMA ask students how teachers respond to correct and incorrect responses.

- PIRLS and TIMSS ask teachers about teaching limitations, including students' nutrition and if they get enough sleep.

- PIRLS and TIMSS ask teachers and principals about the emphasis on academic success, while WEI-SPS asks teachers about school goals and achievement expectations. EGRA and EGMA ask teachers and principals about their expectations of learning levels.

- Assessing and monitoring learning progress, school reports and frequency of tests are addressed in the SACMEQ teacher data collection. Other surveys cover types of formative assessment (LLECE teacher and principal levels); student assessment at classroom level (WEI-SPS teacher level); and monitoring each child's progress (EGRA and EGMA teacher and principal levels).

- Questions about classroom organisation and management, grouping of students and multi-grade instruction are included in PIRLS, TIMSS, PASEC, LLECE and WEI-SPS teacher instruments. The SAQMEC student instrument and the WEI-SPS teacher instrument also ask about personalised learning support and internal differentiation.

- A number of surveys include questions about homework (SACMEQ student, LLECE teacher and principal, EGRA/EGMA student).

- PIRLS and TIMSS ask principals about the evaluation of teacher practice. EGRA and EGMA ask teachers and principals specifically about supervision and classroom visits. WEI-SPS asks principals about professional development.

From the surveys reviewed, a number of domain-related aspects of the quality of education are considered relevant for the context of PISA-D:

- PIRLS and TIMSS ask teachers about reading instruction strategies, assessment practices for reading, use of different reading material, teacher support to develop reading comprehension skills, dealing with reading difficulties, remedial instruction and options for advanced readers, and reading homework. The school-level instruments ask about the emphasis on reading and literacy skills.

- WEI-SPS includes questions about active teaching in reading and mathematics.

- SAQMEQ asks teachers about training for specific subjects and about subject-matter knowledge.

Learning time

Learning time is an important indicator in regard to schooling in developing countries, not only in regard to enrolment, but also in relation to attendance and absence, child labour and class time devoted to the language of instruction, mathematics and science (Willms and Tramonte, 2014: 8, 9). Learning time needs to be captured for learning undertaken both in and out of school (Willms and Tramonte, 2014: 11, 12).

In PISA, information about learning time is collected on student and school levels and covers domain-related learning time (through the student questionnaire), attendance and truancy (through student and school questionnaires), and enrolment and attrition (school questionnaire).

The surveys reviewed also use a number of other in-school indicators of learning time that may be of interest for PISA-D. The LLECE teacher questionnaire notes student

attendance across school shifts (morning, afternoon, intermediate or complete day). The SACMEQ teacher and principal questionnaires measure teaching hours per week. The SAQMEQ principal questionnaire also asks about school days "lost", while the EGRA/EGMA principal questionnaire asks about unofficial school closures during a given year.

PIRLS and TIMSS teacher questionnaires ask about the time students spend on homework.

Indicators of out-of-school factors focus on the impact of child labour on learning time. PASEC and LLECE student questionnaires ask about: types of work (in the household, in agriculture, retail; in/outside the home); amount of work (days per week and hours per day); if students are paid for working; and if working hinders learning, school attendance or causes fatigue during instruction.

Children's participation in the labour force can also be seen as an indicator related to both learning time and to socio-economic status. However, including this topic under socio-economic measures may unnecessarily complicate the measurement of family SES. Variables about children's working can be implemented alongside SES measures (Willms and Tramonte, 2014: 17).

School resources

PISA includes a number of measures in relation to school resources, captured in the school questionnaire. These measures include:

- the size, structure and organisation of the school, including its student and teacher bodies, human resources, and responsibility for specific decision-making

- funding sources

- school resources, including didactic material and facilities, student/computer ratio, school buildings and facilities

- school location, including the size of community.

The PISA 2015 teacher questionnaire will include questions about human resources and teaching conditions, such as teachers' employment status, job experience, workplace selection, subjects studied, and if teachers are teaching in the modal grade for 15-year-old students.

Measures of school resources in developing country contexts need to encompass very low levels of school resources. In many cases, developing countries may lack the resources that would be taken for granted in high-income countries.

Willms and Tramonte distinguish four groups of school resources: material resources, schooling processes, teachers' working conditions and human resources. They posit that PISA-D might usefully extend the "regular" PISA measures of school resources to include a small set of questions relating to material resources. In particular, these questions should focus on basic services, didactic facilities and didactic resources in developing countries (Willms and Tramonte, 2014: 12, 21).

A number of international surveys collect data on basic services, didactic facilities and didactic resources. Relevant questions are concentrated in teacher and principal data collection instruments in SACMEQ, PASEC, EGRA and EGMA, ASER and Uwezo. A fourth category, "other", was added to include topics that are also considered relevant for

PISA-D in regard to school resources – safety at school, teacher satisfaction, staff stability, and funding/grants.

Data collected about basic services relates to:

- school size, including the total number of students in the school's biggest shift (in PASEC and SACMEQ principal questionnaires)

- quality and condition of school buildings (in PASEC and SACMEQ principal questionnaires); with specific items about the cleanliness of the school and surrounds, any major repairs required, and the presence of playgrounds, walls and security guards (in EGRA and EGMA school observations, ASER school facilities observation, and Uwezo principal questionnaire)

- school infrastructure, such as the availability of electricity and telephones, the availability and condition of school resources and school facilities, and whether school facilities are shared between more than one school (various items across PASEC and WEI-SPS principal questionnaires, and EGRA and EGMA school observations and principal questionnaires)

- student/toilet ratio (in the SACMEQ principal questionnaire), the number of functioning toilets in total and for girls (PASEC, EGRA and EGMA school observation, and ASER school facilities observation)

- the presence and functioning condition of a water source (EGRA and EGMA school observation, ASER school facilities observation, and Uwezo principal questionnaire)

- the availability, timing and cost of school meals, the availability of cooking facilities at school (in SACMEQ and ASER principal questionnaires)

- food, transportation, medical and clothing programmes (in LLECE principal questionnaire).

Data collected about didactic facilities is collected across the SACMEQ, PASEC and WEI-SPS teacher questionnaires, the ASER and Uwezo principal questionnaires, and the EGRA and EGMA classroom observations. This data about didactic facilities relates to:

- workspaces, meaning whether there are spaces for students to sit and to write, and where students are seated

- classroom infrastructure, furniture and equipment, such adequate number of seats, adequate lighting in classroom, availability of a blackboard

- classroom resources and teachers' materials, such as boards, chalk, pen, notebook, teacher manuals, teacher lesson plan books and so on.

Data is also collected about didactic resources across many of the reviewed survey instruments. For example, an item about library resources was included in each of the PIRLS and TIMSS teacher questionnaires, PASEC, LLECE and Uwezo principal questionnaires, EGRA/EGMA principal questionnaire and school observation, and the ASER school facilities observation.

Similarly, the availability, quality and frequency of use of pedagogical resources, including teaching resources, educational material, classroom texts and resources for reading instruction were addressed in various items across the PIRLS, TIMSS, SACMEQ and LLECE teacher questionnaires, and the PASEC principal questionnaire.

Surveys also addressed student learning materials, such as:

- whether students own textbooks, and whether students can borrow textbooks from school (in SACMEQ student and principal questionnaires)

- the distribution of textbooks for particular subjects, such as French and Mathematics (PASEC student questionnaire)

- the presence and number of books other than textbooks for reading (EGRA and EGMA classroom observations; EGRA/EGMA, ASER and Uwezo principal questionnaires)

Additional questions relating to didactic resources included:

- the availability or shortage of resources and technology (in the PIRLS and TIMSS principal questionnaires)

- specific resources, such as televisions and photocopiers (LLECE principal questionnaire)

- if computers are available for the use of children (ASER school facilities observation)

- the availability of writing materials, the number of students with pencils, and the display of students' work and instructional material on classroom walls (EGRA and EGMA classroom observations and principal questionnaires).

Data about other material resources that may be relevant to PISA-D relates to:

- safety, school violence and the presence of a security guard (PIRLS, TIMSS, LLECE, EGRA and EGMA teacher and principal questionnaires)

- teacher satisfaction, including the impact of travel distance, if teacher housing is provided and the quality of it, salary levels, quality of educational material, professional development (SACMEQ teacher questionnaire)

- staff stability, in terms of the proportion of teachers at the school for five years or more (WEI-SPS principal questionnaire)

- funding sources (LLECE), school-grant information and repairs, purchases and expenditures (ASER and Uwezo principal questionnaires).

Family and community support

In the context of economically developing countries, family and community support may have an impact on the learning of children living in poverty (Willms and Tramonte, 2014: 8).

Measures of family and community support have been implemented in PISA before, capturing communication with parents, cultural capital, and family involvement. The PISA 2012 school questionnaire included questions about parental expectations towards school and parents' participation in school activities. The parent questionnaire looked at cost of educational services, attitudes to the child's school, parental support for learning in the home and parents' participation in school activities.

Willms and Tramonte (2014: 11, 13) argue that the measures used in PISA should be enhanced in order to create scales that distinguish between parental involvement, social capital and cultural capital, and that are of relevance for PISA-D countries. During the

review of international surveys, relevant questions in regard to family and community support were identified from PIRLS and TIMSS, SACMEQ, LLECE, PASEC, EGRA and EGMA, ASER and Uwezo as well as PIAAC, STEP and LAMP. Questions are included on student/respondent, parent/head of household, teacher and school level.

Data collected about family support relates to tuition, home study support by parents, parents' involvement in the child's education, parents' opinion about the child's school, and other parental support, such as providing a meal for the child before school or having knowledge about performance.

Several questions in the Uwezo instruments seem to address both family and community support, such as parents' sense of how much their opinions about education are heard by local and national officials, parent's views of the most pressing issues facing the community, and parents' awareness about the Uwezo assessment itself.

International survey data collected about community support relates to school community contribution factors and school community problems, community infrastructures and the average income level of the school's immediate area (high, medium or low). These issues are addressed across various items in PIRLS, TIMSS, PASEC, SACMEQ and WEI-SPS principal questionnaires and ASER and Uwezo village observations.

Moreover, PIAAC includes specific measures about cultural capital, household composition and parental home. LAMP looks at human and social capital. Questions cover social context and the literacy levels in the environment, as well as household characteristics and structure, such as the number of individuals living in the household, classified by their relationship to the head of the household, age, sex, and highest level of education.

Health and wellbeing

An eighth category, "health and wellbeing", was added as it was considered relevant for PISA-D during the review of international assessments. Health indicators are considered important for people of all ages because health affects the ability to learn and work. At the same time, the kind of work an individual does affects his or her health status (Pierre, et al., 2014: 21). Also Willms argues that physical and mental health is a key outcome of education, similar to achievement and engagement, and that "health, achievement and engagement affect each other in an interactive process that begins during the primary grades and continues through to adulthood" (Willms and Tramonte, 2014: 5). PIRLS results show that "teachers reported limiting instruction because about one-quarter of the students were suffering from lack of basic nutrition and nearly half from not enough sleep" (Mullis et al., 2012b: 201).

Health and wellbeing factors are covered in several international surveys.

PASEC collects data from students, teachers and principals on wellbeing at school as a factor of the school environment.

Uwezo asks principals about health services such as the presence of a nurse, provision of sanitary items for girls, availability of drinking water and food programmes. It also asks principals to identify the main health issue keeping children out of school, with options of malaria, diarrhoea, cough/flu or other.

PIAAC uses a single item on subjective health: "In general, would you say your health is excellent, very good, good, fair, or poor?" (OECD, n.d.-c: 106).

STEP collects information about a number of key health indicators: height (in centimetres), weight (kilogrammes), level of life satisfaction, existence and kind of health insurance, and number of days the individual was prevented from working during the last four weeks due to sudden illness, accident or chronic illness.

LAMP asks about personal wellbeing and health-related literacy. Respondents are asked about their health condition and if they can perform basic functions like filling in medical forms, reading medical labels and food labels.

Implications

The range of important measures is quite extensive. PISA-D contextual questionnaires should be highly focused in order to accommodate the limited time and to reduce the burden on respondents, and particularly given the likely relatively low reading ability of the intended target population of PISA-D (students as well as parents).

As outlined in the terms of reference for PISA-D, the student and school questionnaires should address policy issues of interest to participating countries, and take about 30 to 35 minutes to be completed (OECD, 2014a). On the student level, PISA-D should collect information with two components:

- a core component with basic demographic information, with key questions from the previous PISA cycles for PISA-D

- a focused component through which in-depth information on one or more specific policy issues identified by the participating countries is collected.

The focused component should be designed specifically for PISA-D in order to address policy issues of interest to the participating countries as per the themes identified in this section.

Early learning opportunities

The PIRLS and TIMSS Learning to Read Survey (2011) may be a useful component for the PISA-D parent questionnaire. It includes information about language spoken in the home, preschool experiences, homework activities, home-school involvement, books in the home, and parents' education and occupation. In addition, this questionnaire collects information on early literacy and numeracy activities, reading and quantitative readiness, and parents' reading activities and attitudes toward reading. Together with information collected from the students, parents' responses will provide a more complete picture of an important context for learning to read and numeracy. The questionnaire is designed to take 10 to 15 minutes to complete (Martin, Mullis and Foy, 2013: 67, 68). Depending on whether or not a parent questionnaire will be implemented in PISA-D countries, it will be worth considering if some of the questions about early reading and numeracy could be included in the student questionnaire. If a parent questionnaire is not considered, we recommended implementing the respective questions on student level to find out whether they can be reliably answered by 15-year olds in PISA-D. Additionally, existing parent questionnaires could be used in the field trial to compare student and parent responses to questions about early learning opportunities.

Questions from LLECE (SERCE Grade 3 and 6 student questionnaires and parent questionnaire) about early reading and how often someone at home reads aloud to the child may also be of interest to PISA-D.

Of particular interest to PISA-D may be questions about the out-of-school status of 15-year-olds, as implemented in ASER and Uwezo. This should be considered if a (household-based) component to reach out-of-school-children is introduced in PISA-D.

Language of home and school

In regard to language at home and in school, it may be worth including language-related questions from the PISA 2012 educational career questionnaire. Of particular interest would be the questions about the first language learned at home, age when test language was learned, the language usually spoken with different groups of people such as parents and friends, and the language used for different activities. These questions would reveal a broader picture of students' familiarity with the test language, which is important (Willms and Tramonte, 2014: 11). In addition, questions about the frequency of speaking the test language at home and the language spoken by the student before school enrolment (both from PIRLS and TIMSS), would be worth including. The PIRLS and TIMSS approach of asking if the books at home are mainly in the test language is also relevant for PISA-D.

Questions used in STEP and LAMP also have potential for PISA-D in providing a full picture of the languages at home, and differentiating between languages that respondents speak, read and write.

One option for the PISA-D teacher questionnaire is to include a question about the languages spoken by the teacher, as for example in PASEC. This would show any correlations between the language of instruction and the language spoken at home at the teacher level. Additionally, teachers could be asked to estimate the number of students that have difficulties understanding the spoken test language (as is done in PIRLS and TIMSS).

PISA-D should also include broader questions about the language of instruction (Willms and Tramonte, 2014: 11). Questions from PISA 2009, PIRLS and TIMSS that may be relevant for PISA-D relate to the proportion of students that have a first or native language, or mother tongue, that is not the test language. The availability of additional instruction for students with a first language other than the test language may also be an issue.

At the school level, questions from LLECE about the language(s) of instruction and indigenous language services and resources may also be of interest to PISA-D. It would be useful for PISA-D to ask about the official time used for teaching the language of instruction, as in WEI-SPS, and about languages in which textbooks are provided, as in Uwezo.

Socio-economic status

PISA-D requires a combined approach of extending the current indicators of the PISA index of economic, social and cultural status to include items at the lower end of the socio-economic scale, and developing new poverty-related measures. The review of international surveys provides valuable information for both options.

Indicators relevant to children living in poverty are included in SACMEQ, PASEC, LLECE, EGRA and EGMA, ASER and Uwezo, and can be summarised as follows:

- parental education, such as if parents can read and write; if parents attended school and the status of their completed education; if parents never attended

school (additional questions from LAMP and STEP about formal and non-formal education may be useful to fully capture parent's education)

- main source of the income and occupation, with options to include unemployed, wage employee (office), transfers (from other people), farming or animal production, wage employee (casual labour), home maker, own business, other

- home facilities, in terms of the structural features of the dwelling, including electricity, water, construction material of the home, type of house, availability of and type of toilet, lighting, method for cooking, water source for washing, number of meals per day

- home possessions, in terms of material possessions in the household as well as personal material possessions of the respondent, including radio, TV, (mobile) phone, computer, internet, cattle, donkeys, camels, sheep or goats, bicycle, motorbike, cart

- educational resources in the household, including educational materials, number of books, reading material (books and newspapers), and whether anyone in the household knows how to use a computer.

STEP, LAMP, ASER and Uwezo, all implemented in developing country contexts, provide a well-established pool of variables for household characteristics. PISA-D can draw from this pool, using relevant variables to extend the index of economic, social and cultural status scale, as well as to develop new poverty-related measures. In STEP, a particular asset index (Pierre, et al., 2014: 15) was created, that bears potential for PISA-D. The STEP asset index is further discussed in the examination of socio-economic and poverty-related measures in section 4.4.2).

Employment information as captured in LAMP, STEP or PIAAC may be of interest for the PISA parent questionnaire, in regard to extending existing measures of the parents' employment status. For example STEP module 4 obtains basic employment information, such as the labour force status (employed, unemployed or inactive; including self-employed – with and without pay; underemployed or holding low-productivity jobs).

Quality of instruction

The international surveys reviewed offer a wide range of factors indicating quality of instruction – both in general as well as domain-related – at the student, teacher and school levels.

PISA measures of general aspects of quality of instruction could be extended for use in PISA-D. Of particular interest are specific pedagogical practices, limitations of teaching (including students nutrition and if they get enough sleep), emphasis on academic success and achievement expectations, assessing and monitoring learning progress, classroom organisation and management (such as multi-grade instruction, grouping of students and personalised learning), homework, evaluation of teacher practice and professional development.

Domain-related aspects of quality of instruction that are of particular relevance for PISA-D are reading instruction strategies, including options for advanced readers and students dealing with reading difficulties, and teacher training for specific subjects, including teachers' subject matter knowledge.

It is important to consider whether PISA-D could sufficiently address these aspects at the school and student level, or whether a teacher questionnaire is necessary. The comparison of international surveys shows that some of the relevant factors are currently collected on both classroom teacher and school principal questionnaires.

Learning time

Learning-time factors include enrolment and school attendance of both students and teachers, as well as time for instruction. These topics are covered in the PISA questionnaire on student and school levels.

PISA-D should include measures of learning-time factors specific to developing countries, especially the impact of child labour. Questions about working outside school are captured in PASEC and LLECE student questionnaires, which ask about the type of work (in the household, in agriculture, retail; inside/outside the home); if the students are paid for working; the amount of work (days per week and hours per day); and if working hinders learning or school attendance, or causes fatigue during instruction.

School resources

Wills and Tramonte (2014: 12, 21) suggest including a small set of questions in PISA-D relating to material resources, focusing on basic services, didactic facilities and didactic resources. Relevant factors were mainly found in those surveys addressing student or child populations in economically developing countries. Relevant questions were found in SACMEQ, PASEC, EGRA and EGMA, ASER and Uwezo.

Factors relating to basic services mainly include conditions of the school building and school infrastructure such as the availability of electricity, toilets, drinking water sources and provision of school meals, transportation and medical and clothing programmes. The main informant for questions related to basic services is the principal. Information is also captured through school observation.

Factors relating to didactic facilities include information about teachers' workspace, classroom resources and infrastructure such as tables, chairs and other furniture, blackboard, chalk, pen, notebook, and adequate lighting in classroom. Main informants for didactic facilities are students and teachers, but also principals and classroom observation.

Factors relating to didactic resources cover teaching resources such as television, photocopier, or computer, availability and quality of educational material, availability of a library, student learning materials such as textbooks, pencils and other writing materials. Quality and frequency of use is mainly captured through the teacher; students are asked about use and ownership of material. For library resources the principal seems to be the main informant.

Other relevant topics that have been identified during the review of international surveys and that are of relevance for PISA-D are school safety, teacher satisfaction (including factors such as travel distance, if teacher housing is provided, or level of salary), staff stability, and issues regarding funding and grants. In respect to receipt and spending of grants the ASER Centre indicates that these items have become more and more detailed over the years. The rationale behind this is that it is important to have information about allocation of resources to the right activities, people responsible for decision-making, flow of funds and if the money reaches where it is supposed to (ASER Centre, 2014: 12-13). School safety, staff stability and issues of funding have mainly

been addressed at school level, whereas information about teacher satisfaction has been captured from the teachers.

Family and community support

Information about parental involvement is captured at all levels – student, parent, teacher and school level. Surveys that include relevant factors for parents' involvement for PISA-D are PIRLS and TIMSS, SACMEQ, LLECE, WEI-SPS and EGRA and EGMA.

Information about community support is mainly captured through the principal. Useful factors and variables can be found in SACMEQ, WEI-SPS, PIRLS and TIMSS and PASEC.

Specific measures of cultural and social capital, which are of relevance for PISA-D, are included in PIAAC and LAMP.

Health and wellbeing

Health and wellbeing are considered important outcomes of education and are of particular relevance to economically developing countries, where students have to deal with malnutrition and the availability of basic health services cannot be assumed. Factors measuring health and wellbeing should therefore be included in the PISA-D context questionnaires. The relevant information can be properly addressed at the student and school level.

A number of international surveys measure factors about health and wellbeing that may be of particular interest for the inclusion in PISA-D.

Uwezo asks principals about health services such as the presence of a nurse, provision of sanitary items for girls, availability of drinking water and food programmes. It also asks principals to identify the main health issue keeping children out of school, with options of malaria, diarrhoea, cough/flu or other.

LAMP asks about personal wellbeing and health-related literacy. Respondents are asked about their health and if they can perform basic functions like filling in medical forms, reading medical labels and food labels.

A health-related literacy component may be of interest to PISA in general, for example as an accompanying questionnaire option in the context of the science literacy assessment.

Technical aspects of contextual data collection instruments

Question formats

In regard to question formats, PISA uses Likert scale (a method of ascribing quantitative value to qualitative data, to make it amenable to statistical analysis) and open response questions. Since the PISA 2012 assessment, formats such as "forced choice", "situational judgement tests", "overclaiming techniques" and "anchoring vignettes" have been introduced and are discussed below.

In the surveys reviewed, across all contextual data collection instruments, the following question formats were used:

- dichotomous questions, mostly with yes/no responses, used particularly in ASER and Uwezo

- nominal variables

- Likert scales, including three, four, five and ten-point scales

- open-ended questions

- rankings.

An example of rankings is found in an Uwezo household survey item about major issues facing the community. The respondent is asked to choose three of nine options and rank the three chosen options in order of importance.

Open-ended questions, which were largely used in ASER and Uwezo, are not very cost or time-effective for data capture, analyses and aggregation, and grouping of information.

Scaling of contextual constructs

Table D.8 in Annex D provides an overview of scaling methodologies applied in the different international surveys for contextual constructs. The right column describes context constructs relevant to PISA-D. Socio-economic measures (as one of the seven priorities) are described in Table D.9.

Scaling/computing of relevant contextual constructs

In PISA two kinds of indices are created from context questionnaire constructs. Simple indices are constructed through arithmetic transformation or recoding. Scale indices are constructed through scaling of multiple items, using a weighted likelihood estimate, and in most cases using a one-parameter item response model (a partial credit model was used in the case of items with more than two categories) (OECD, 2014b).

For scale indices, in general, the scaling is done in three stages. First, the item parameters are estimated from equal-sized subsamples of students from all participating countries and economies. Second, the estimates are computed for all students and all schools by anchoring the item parameters obtained in the preceding step. Third, the indices are then standardised so that the mean of the index value for the OECD student population is 0 and the standard deviation is 1 (with countries being given equal weight in the standardisation process) (OECD, 2014c: 260).

The combination of item response theory scaling methodology and computation of simple indices is commonly used in large-scale international studies, as well as in the household-based studies PIAAC and LAMP.

In STEP, the mostly simple indices are derived from Likert scales (Pierre et al., 2014: 69).

In PIRLS and TIMSS each context scale (derived from item response theory scaling) was divided into regions, corresponding to high, middle and low values on the construct. The cutpoints between the regions were defined in terms of response categories to facilitate interpretation of the regions (Martin et al., 2012).

No methodological guidelines for processing of contextual constructs are provided in EGRA and EGMA, ASER and Uwezo. Usually some simple computations are carried out to aggregate or average variables, or to create ratios (such as teacher/student ratio).

Relevant context constructs from international surveys

The following context constructs used in international surveys are of particular relevance for PISA-D. These constructs are organised according to the seven key areas of focus identified by Willms and Tramonte (see section 4.3).

Early learning opportunities

Children's early literacy activities before beginning primary school are measured in PIRLS. The scale is based on the parent's report of how often they do nine activities, such as reading books, telling stories, singing songs and playing word games.

Children's early numeracy activities before beginning primary school are measured in TIMSS. The scale is based on the parent's report of how often they do six activities, such as saying counting rhymes or singing counting songs, counting different things, playing with building blocks or construction toys.

PIRLS also collects data on whether children could do early literacy tasks at the beginning of primary school. The scale is based on parents' responses to how well their children could do five tasks, including recognising most of the letters of the alphabet, reading some words, reading some sentences.

TIMSS considers whether children could do early numeracy tasks at the beginning of primary school. The scale is based on parents' responses to six statements, such as whether children count independently or recognise different shapes.

Language of home and instruction

PISA's language background construct indicates whether a students' language at home is the same as the language of assessment or a different language than the language of assessment.

Quality of instruction

PIRLS and TIMSS principal and teacher report constructs include a scale of school emphasis on academic success. This scale considers five aspects, including teachers' understanding of the school's curricular goals and teachers' expectations for student achievement.

PIRLS includes a scale of emphasis in early grades on reading skills and strategies, based on principals' responses about the earliest grade at which each of eleven reading skills and strategies were emphasised.

PIRLS and TIMSS include a scale of collaboration to improve teaching. The construct is based on teachers' responses to how often they interacted with other teachers in each of five teaching areas. The areas include discussing how to teach a particular topic and visiting another classroom to learn more about teaching.

PIRLS and TIMSS include a scale on instructions to engage students in learning. The construct is based on teachers' responses to how often they used each of six instructional practices. Practices listed include summarising what students should have learned from the lesson and praising students for good effort.

LLECE includes an index of educational opportunity. The construct is based on measures of classroom time, learning resources, school library resources, financial resources, school infrastructure, and teacher and leader quality. The index also considers

processes that mediate pedagogy, such as curriculum coverage, language of instruction, school autonomy, use of teaching materials, homework and school climate. Analyses are conducted at the classroom, school and education system levels.

School resources

PIRLS and TIMSS include a number of constructs about school resources.

They include scales of the extent to which instruction is affected by resource shortages (in reading and mathematics respectively). The constructs are based on principals' responses concerning the availability of general and subject-specific resources in the school and classroom.

They also include a teachers' working conditions scale. The construct is based on teachers' responses concerning five potential problem areas: school buildings needing significant repair; classrooms being overcrowded; teachers having too many teaching hours; teachers not having adequate workspace; and teachers not having adequate instructional materials and supplies.

A scale for the safety and order of the school is based on teachers' degree of agreement with five statements. Statements include: this school is located in a safe neighbourhood; I feel safe at this school; and the students behave in an orderly manner.

The principal questionnaire includes a school discipline and safety scale. The scale is based on principals' responses concerning ten potential school problems, including students arriving late at school, unjustified absenteeism, vandalism, and so on.

LLECE includes an index of accessibility of basic school services in the principal (census) questionnaire. The construct is based on five items requiring yes/no answers, if the following exists in the school: electricity/lights; drinkable water; sewage system; phone; sufficient number of bathrooms.

The LLECE index of school infrastructure is based on 15 items the principal questionnaire. The items ask whether the school has: a principal's office; additional offices (secretary/administration); staff room; sports field/court/oval; science room; gym; school garden; computer room; auditorium; kitchen' cafeteria; art/music room; medical office; speech-psychology services; school library.

Family and community support

SACMEQ includes a school community contribution factor. The construct is based on the sum of the presence of community contributions towards nine school activities. Activities include construction and maintenance of school buildings; construction and repair of school furniture; provision of school meals; buying of textbooks, stationery and supplies; payment of teacher salaries; and extra-curriculum activities.

Implications

Question formats

PISA questionnaires include a number of self-reported measures. These include motivation, self-concept, engagement and enjoyment. Analyses of these measures often show a correlation between performance and attitudes (for example, interest in mathematics and mathematics performance). There are also concerns about the cross-cultural comparability of self-reported measures. However, these can be addressed

– in PISA the scales that were adjusted for differences in response behaviours by means of anchoring vignettes were shown to have a positive correlation with mathematics performance. Anchoring vignettes provide a comparatively inexpensive way of creating an anchor within the survey context itself. The idea is to compare respondents' self-assessments to the respondents' assessments of hypothetical people described in short vignettes that have known characteristics, and to use the latter to adjust the former.

It will certainly be of interest for PISA-D to include item formats that allow for an adjustment of self-reported measures to further explore and potentially increase cross-country comparability. We recommend undertaking analyses to examine the extent of different patterns of response styles in the countries participating in PISA-D.

Scaling of contextual constructs

We recommend that PISA-D follow the procedures used in PISA to scale context questionnaire scales. This includes employing item response theory scaling methodology (for example, see OECD, 2009: 7-9). This scaling technique is robust for comparisons across different samples and over time.

PIRLS and TIMSS context questionnaire scaling could be of particular interest for PISA-D. Given that PIRLS, TIMSS and PISA have all used ConQuest item response modelling software, the algorithm underlying this particular scaling would probably be similar across these assessments.

Questionnaire scales of relevance for PISA-D are mainly from PIRLS and TIMSS. Of particular interest are the scales for early literacy and numeracy activities before beginning primary school and for early literacy tasks at the beginning of primary school (early learning opportunities), as well as the scale for instructions to engage students in learning (quality of instruction).

Relevant indices identified from LLECE are educational opportunity, relating to learning time, learning resources, school resources and infrastructure, quality of instruction; accessibility of basic school services and school infrastructure.

The SACMEQ school community contribution factor could also be valuable for PISA-D.

Socio-economic and poverty-related measures

As mentioned earlier, it is envisaged that PISA-D will extend the current indicators of the PISA index of economic, social and cultural status to include items at the lower end of the socio-economic scale, as well as develop new poverty-related measures.

PISA uses a number of SES-related measures. The measure for parents' occupational status involves recoding ISCO codes into International Socio-Economic Index (ISEI) occupational status codes. The measure for parents' educational level has used the 1997 version of UNESCO's International Standard Classification of Education (ISCED 97), but will use ISCED 11 for PISA 2015. Other measures include wealth (based on home possessions), home educational resources (including books at home) and cultural possessions. PISA's index of economic, social and cultural status (ESCS) is derived from indices of all these measures (for details see Table in Annex D).

International surveys do include indicators relevant to children living in poverty, but do not measure them distinctly from socio-economic status.

Home resources, possessions and assets

Characteristically, measures that include indicators relevant to children living in poverty are mainly based on home resources, characteristics of the household, and possessions and assets. Surveys that are of interest to PISA-D are SACMEQ, LLECE, EGRA and EGMA, ASER and Uwezo, as well as STEP and LAMP. For details of the SES-related scales see Table in Annex D. For an interesting discussion of SACMEQ's poverty-related measures, see Dolata (2005).

In SACMEQ a student socio-economic status factor is derived from 18 items. Items about home possessions ask whether the household has books, newspapers, magazines, radio, television, VCR, cassette player, telephone, refrigerator, car, piped water, a table to write on. Parental education items cover mother's education and father's education. Other items include home quality (floor, roof, outside walls) and lighting to read (Dolata, 2005: 40).

The household resources measures in EGRA/EGMA, ASER and Uwezo are very similar. They include variables such as type of house, electricity connection, availability of toilet, type of toilet, method for cooking food, presence of a water source, number of meals per day. Household possessions in ASER and Uwezo also include availability of possessions such as a radio, TV, mobile phone and reading material (books and daily newspapers); cattle, donkeys, camels, sheep/goats, bicycle, motorbike, cart. In addition, ASER and Uwezo collect data on parents' educational background; ASER asks if anyone in the household knows how to use a computer, attended school and status of completed education, never attended school, and if mother and father can read.

For the Uwezo regional report an SES indicator was created. Households in the survey are categorised into three socio-economic groups according to durable assets owned, access to electricity and/or clean water, and mother's formal education level (Uwezo, 2014: 16). Children are then categorised into three groups: non-poor, poor and ultra-poor.

Similar indices are created in PASEC and LAMP. PASEC uses standard of living categories of poor, intermediate and rich. In LAMP respondents are classified into four socio-economic groups – affluent (well-off), comfortable, poor or subsistence level – based on the structure of the household and the available equipment.

Interesting for PISA-D is the asset index constructed in STEP for urban areas, based on the information on dwelling characteristics and household assets (see section 3.1.1). The asset index is used as a proxy for wealth. Since the focus of the survey is to obtain detailed information at the individual level, the household-level information is kept to a minimum (Pierre, et al., 2014: 14).

This STEP asset index was constructed using factor analysis over a set of indicator variables for the different types of assets and dwelling characteristics (Pierre, et al., 2014: 15). All national-level estimations were weighted using each country's sample weights, "in order to reflect underlying measures of welfare" (Pierre, et al., 2014: 15). Therefore, during the selection of the variables, variables with extremely skewed distributions (with means across assets and dwelling characteristics below 0.02 and above 0.98) were excluded from the analysis. Deliberation was made over the inclusion of agricultural assets, which were considered productive assets and not an indication of wealth per se (Pierre, et al., 2014: 15). Moreover, variables with low factor loading (less than 0.1) on the un-rotated first factor of the overall asset index were excluded for the final asset index (Pierre, et al., 2014: 15).

The asset index itself was constructed on a country-by-country basis according to the following process (Pierre, et al., 2014: 15):

1. An indicator variable was created for each of the dwelling characteristics and assets available in Module 1b of the STEP household questionnaire.

2. The variables that did not comply with the first selection criteria were dropped.

3. An overall asset index was generated using factor analysis and it included all the available asset and dwelling-related variables. In this stage, the factors with an Eigen value of more than 1 were selected. Eigen value is a scalar associated with a given linear transformation of a vector space and having the property that there is some nonzero vector which when multiplied by the scalar is equal to the vector obtained by letting the transformation operate on the vector; especially : a root of the characteristic equation of a matrix

4. A varimax rotation is used to simplify the expression of a particular sub-space in terms of just a few major items each and was employed using the selected factors from the previous step.

5. A Cronbach's alpha (or scale reliability coefficient) was estimated for this overall asset index.

6. Indexes for each domain (dwelling characteristics, primary assets, and secondary assets) were constructed by following the same procedure from steps 3 to 5.

7. A pairwise correlation was estimated for each of the domain indexes compared to the overall asset index to determine the level of association.

8. Variables that did not meet the third selection criteria were dropped.

9. A final asset index was constructed based on the factors with an Eigen value of more than 1.

Given that assets play an important role in regards to poverty-related measures, the asset index created in STEP may be a valuable resource for PISA-D.

In LLECE an index of socio-economic and cultural background is created, which includes children's wellbeing and cultural access at local, regional and global levels. The index also emphasises home assets, assuming that these facilitate access to culture and learning. LLECE also includes an index of educational home environment, which considers parental involvement in education as well as current and early childhood education.

School and classroom resources

In addition to home resources, possessions and assets, school and classroom resources are also related to socio-economic and poverty-related measures.

SACMEQ includes both a classroom resources factor and a school resources factor. The classroom resource factor is computed from the sum of the existence of eight items in the classroom: writing board, chalk/marker, wall chart, cupboard, bookshelves, classroom library or book corner, teacher table, and teacher chair. The school resources factor was computed in two ways. The first is a sum of the existence of 22 school resource items in the school including a school library, school meeting hall, staff room, separate office for school head, sports area, water, electricity, telephone, fax machine, overhead projector, radio, TV set, photocopier and computer. The second way calculates a Rasch score

involving school resources items as well as classroom resource items, such as teacher table, teacher chair, sitting places, cupboard and bookshelves (Hungi, 2011a).

WEI-SPS uses indices of social advantage that are of relevance for PISA-D. The index of social advantage of school intake is based on principals' responses about the number of students (none, most, all) whose parents are educated and the number who receive food or clothing programmes, and on the SES of the school intake compared to national GDP per capita. The social advantage of classroom intake index has been computed based on teacher's responses on the number of students (none, most, all) who undertake child labour or who have family health problems, among other issues (UIS, 2009a: 70, Appendix III).

Implications

Willms and Tramonte (2014) concluded that the current PISA measure of socio-economic status does not include a sufficient number of items at the lower end of the scale to adequately describe the populations of students in the PISA-D countries (OECD, 2014: 56-57). The expert paper presents two options for addressing this issue:

1. extending the current indictors of the PISA index of economic, social and cultural status to include more items at the lower end of the SES scale

2. developing new poverty-related measures.

Willms and Tramonte recommend pursuing the first option as a starting point. This may have limitations, in that the resulting scale will not be uni-dimensional and the new items will differ in their relationship to achievement for low and high-SES students. The authors argue that a combined approach could be pursued, with attention to the goals of international comparability and maintaining a link to the current PISA framework (OECD, 2014a: 56-57).

A measure of socio-economic status for PISA-D should be:

• a reliable and valid measurement of SES within each country

• a tool for accurate assessment of low levels of SES and poverty within each country and across countries

• a comparable measure of SES and its variability across the participating countries (OECD, 2014a: 56-57).

The review of international surveys shows that SES-related measures applied in international surveys conducted in developing country contexts commonly include indicators relevant to children living in poverty, but do not measure them distinctly from SES. Such indicators tend to be mainly based on home resources, characteristics of the household and possessions and assets. A good source for factors related to household resources and possessions are EGRA/EGMA, ASER, Uwezo and LAMP. The factors can be used to categorise responses from households and children, as for example in Uwezo, PASEC or LAMP.

The asset index created for STEP is based on information on dwelling characteristics and household assets (Pierre, et al., 2014: 15). Despite being at the lower end of a global SES range, there is a breath of levels of economic development within and across the countries participating in PISA-D. The challenge in creating an asset index for PISA-D would be to find assets that function as indicators to differentiate meaningfully between different levels of SES equally well across all countries.

A number of indices from other international surveys are considered particularly relevant for PISA-D:

- the student socio-economic status factor computed in SACMEQ (Dolata, 2005: 40) based on measures of home possessions, home quality and parental education

- the index of socio-economic and cultural background created in LLECE, which includes home assets, but also children's wellbeing and cultural access at local, regional and global levels

- the index of educational home environment created in LLECE, which considers parental involvement in education as well as current and early childhood education.

In addition to home resources, school and classroom resources can also be SES indicators. Relevant for PISA-D are the classroom resource factor and school resource factor created in SACMEQ, and the social advantage of school intake index and social advantage of classroom intake index created for WEI-SPS.

Poverty-related measures often ask if respondents have a particular resource, such as textbooks or a television. Meaningful results can also be gained by also asking whether the respondents would actually like to have an item they do not own. In other words, the response options would be: I have this; I do not have this but would like it; and I do not have this and I do not want or need it. Such a response scale could be explored further in the context of PISA-D.

There is a need for PISA-D to capture different countries' experiences with their own variables for measuring socio-economic status. Countries participating in PISA-D have a history of data collection and valuable experience on how to effectively assess socio-economic status in their specific cultural and geographical contexts.

PISA-D should look for options to ensure cross-cultural comparison. Three aspects are crucial:

- translating, adapting and verifying

- constructing context indices

- data analyses.

Cross-country and cultural comparability greatly depend on translation and adaptation procedures, including standardisation and verification. Country involvement is vital, and facilitation by national centres, project managers and experts. Country involvement includes reviewing context questionnaire frameworks and questionnaire items. This process ensures the face-validity (or face value) and cultural appropriateness of the content, and reduces the potential for translation issues.

PISA has well established procedures for translating, adapting and verifying questionnaire materials which should also be followed in PISA-D to ensure the rigour of this part of the assessment. Still, it may be worth incorporating translatability assessments to reduce the cost associated with this process (cApStAn, 2015).

Cross-country comparability also depends on the construction of context indices. PISA-D should include anchoring vignettes (as described above) and overclaiming techniques (these require respondents to rate their familiarity with a list of general knowledge items, such as persons, places, things). PISA-D should also explore further the application of forced choice content and format. This format was only partly pursued in

PISA 2012 due to the ethical considerations of forcing students to make a choice where, in reality, such a choice would not have to occur: for example, forcing students to theoretically choose between a career in science or mathematics when, in reality, careers in those areas frequently involve both mathematics and science. PISA-D should also explore further the situational judgement test format, which was developed for the problem-solving approaches in PISA 2012, but ultimately found to have unsatisfactory levels of reliability. These formats may be productively pursued through development and cognitive testing of such items in PISA-D countries.

Finally, PISA-D should ensure data analyses after field trialling and after the main study focus on cross-country comparability. Data analyses need to capture the validity of questionnaire items across countries and to check that items work in the same way in all countries. This is essential for cognitive as well as contextual items. It is essential that countries review these analyses. Data adjudication ensures that data are valid, reliable and objective. This is done by all international large-scale assessments. PISA has already established highly elaborate standards in this regard. PISA has introduced measures to adjust relevant context questionnaire scales (self-reported measures) to ensure cross-cultural comparability. This hasn't been found in any of the other surveys reviewed.

PIRLS and TIMSS do provide evidence that the context questionnaire scales provide comparable measurement across countries. These surveys compute reliability coefficients for each scale for every country and benchmarking participant. A principal components analysis of the scale items is conducted (Martin et al., 2012: 6). This analysis looks for a positive relationship between indicators of an effective learning environment and indicators of achievement. A strong correlation, across all countries, is seen as evidence of the validity of the context questionnaire scales (Martin et al., 2012: 9)

As has been done in the major international large-scale assessments, analyses should be aimed at examining the extent to which scales, and hence the constructs they intend to measure, have consistent dimensionality and validity across participating countries. This should be done particularly at the field trial stage to ensure that the most valid measures will be selected for the main study.

Confirmatory factor analyses and multi-group confirmatory analyses can be used to examine the dimensionality of scales within and across countries. While it cannot be expected that correlations between the same constructs are exactly the same across countries, a similarity of patterns could be expected.
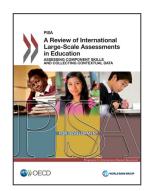
Notes

1. In addition to contextual data at student, classroom and school level, data at the system level play an important role within PISA, and will also be of particular importance for PISA-D. System-level data collection in PISA is conducted through OECD NESLI (INES Network for the Collection and Adjudication of System-Level Descriptive Information on Educational Structures, Policies and Practices). For PISA-D, a separate paper on "System-Level Data Collection" has been jointly commissioned by the OECD and the World Bank from the UIS to investigate the current status of system-level data collection and availability of participating countries in PISA-D.

References

Allen, J. et al. (2013), "The development of the PIAAC background questionnaires" in *Technical Report of the Survey of Adult Skills (PIAAC)*, OECD, Paris.

ASER Centre (2014), *Annual Status of Education Report (Rural) 2013*, ASER Centre, New Delhi.

ASER Centre (n.d.), *How Far has India Come in Guaranteeing Education? The Right to Education Act and ASER Findings 2010-2012*, ASER Centre, New Delhi, http://img.asercentre.org/docs/Publications/aser6pageronstatusofrteimplementation_30 thmar2013.pdf.

Banerji, R. and S. Bobde (2013), "Evolution of the ASER English Tool" in V. Berry (ed.), *English Impact Report: Investigating English Language Learning Outcomes at the Primary School Level in Rural India*, British Council, London.

cApStAn (2015), "cApStAn Linguistic Quality Control, Inc.", www.capstan.be/content/tr anslatability_assessment.html (accessed 23 May 2015).

Carroll, J. (1963), "A model of school learning", *Teachers College Record* 64(4), Teachers College, Columbia University, New York, pp.722-723.

CONFEMEN (2013), *CONFEMEN Programme for the Analysis of Education Systems: PASEC*, CONFEMEN, Dakar.

CONFEMEN (2012), *Améliorer la Qualité de l'Education au Tchad : Quels sont les Facteurs de Réussite? Évaluation Diagnostique PASEC-CONFEMEN 2e et 5e du Primaire Année Scolaire 2009/2010*, CONFEMEN, Dakar.

Creemers, B.P.M. (1994), *The Effective Classroom*, Cassell, London.

Crouch, L. (2009), *The Snapshot of School Management Effectiveness: Report on Pilot Applications*, RTI International, North Carolina.

Crouch, L. (2008), "Snapshot of school management effectiveness aims, initial development, instruments, methods", presentation, SSME workshop, Washington DC, 18 December 2008, www.eddataglobal.org/management/index.cfm?fuseaction=pubDe tailandID=164.

Dolata, S. (2005), "Construction and validation of pupil socio-economic status index for SACMEQ education systems", conference paper, International Invitational Educational Policy Research Conference, Paris, 28 September to 2 October 2005.

Foy, P., A. Arora and G.M. Stanco (2013), *TIMSS 2011 User Guide for the International Database, Supplement 2: National Adaptations of International Background Questionnaires*, Lynch School of Education, Boston College, Massachusetts.

Hooper, M., I.V.S. Mullis and M.O. Martin (2013), "PIRLS 2016 context questionnaire framework", in I.V.S. Mullis and M.O. Martin (eds.), *PIRLS 2016 Assessment*

Framework, TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement, Massachusetts, pp. 33-55.

Howie, S. et al. (2012), *PIRLS 2011: South African Children's Reading Literacy Achievement, Summary Report*, Centre for Evaluation and Assessment, University of Pretoria, Pretoria, www.up.ac.za/media/shared/Legacy/sitefiles/file/publications/2013/pirls_2011_report_12_dec.pdf.

Hungi, N. (2011a), *Accounting for Variations in the Quality of Primary School Education*, SACMEQ, Paris, www.sacmeq.org/?q=publications.

Hungi, N. (2011b), "Characteristics of Grade 6 pupils, their homes and learning environments", *SACMEQ Working Paper*, SACMEQ, Paris.

LLECE (2009), *SERCE: Segundo Estudio Regional Comparativo y Explicativo: Los Aprendizajes de los Estudiantes de América Latina y el Caribe; Reporte Técnico*, (Second International Comparative Study of Student Learning in Latin American and the Caribbean: Technical Report), Office Santiago and Regional Bureau for Education in Latin America and the Caribbean, LLECE, Santiago.

Martin, M.O. et al. (2012), "Creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales", 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.

Martin, M.O., I.V.S. Mullis and P. Foy (2013), "PIRLS 2016 assessment design and specifications", in I. V. S. Mullis and M. O. Martin (eds.) *PIRLS 2016 Assessment Frameworks,* TIMSS and PIRLS International Study Center and IEA, Chestnut Hill, MA and Amsterdam, pp. 57-69.

Messaoud-Galusi, S. et al. (2012), *Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management in Doukkala Abda, Morocco*, RTI International, North Carolina.

Mullis, I.V.S. et al. (2012a), "Assessment framework and instrument development", 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.

Mullis, I.V.S., et al. (2012b), *PIRLS 2011 International Results in Reading*, TIMSS and PIRLS International Study Center and IEA, Chestnut Hill, MA and Amsterdam.

OECD (2014a), *Call for Tender 100000990 - PISA for Development Strand A and Strand B*, OECD, Paris, http://tinyurl.com/prjry24.

OECD (2014b), *PISA 2012 Technical Report*, OECD Publishing, Paris, www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

OECD (2014c), *PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science,* OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264208780-en.

OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy,* OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264190511-en.

OECD (2013b), *The Survey of Adult Skills: Reader's Companion*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264204027-en.

OECD (2009), *PISA 2006 Technical Report*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264048096-en.

OECD (n.d.-a), "PISA 2015 draft questionnaire framework", www.oecd.org/pisa/pisa products/PISA-2015-draft-questionnaire-framework.pdf (accessed 5 August 2015).

OECD (n.d.-b), "OECD Skills Surveys: PIAAC: Main elements of the survey", www.oecd.org/site/piaac/mainelementsofthesurveyofadultskills.htm (accessed (20 November 2014).

OECD (n.d.-c), "PIAAC Background Questionnaire: MS version 2.1 d.d. 15-12-2010," OECD, Paris.

Pierre, G. et al. (2014), *STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills*, working paper, World Bank Human Development Network, Washington DC.

Purves, A. C. (1987), "*The evolution of the IEA:* A memoir", *Comparative Education Review* 31(1), University of Chicago Press, Chicago, pp. 10–28.

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 (2006), *Literacy Assessment and Monitoring Programme (LAMP) Background Questionnaire (BQ)*, UNESCO Institute for Statistics, Montreal.

Uwezo (2014), *Are Our Children Learning? Literacy and Numeracy across East Africa 2013*, Uwezo and Hivos/Twaweza, Nairobi.

Uwezo (2013), *Uwezo 2013 Annual Plan and Budget*, Uwezo, Nairobi.

Willms, J.D. and L. Tramonte (2014), "Towards the development of contextual questionnaires for the PISA for development study", *OECD Education Working Papers*, No. 118, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js1kv8crsjf-en.

Yu, A., and D. Ebbs (2012), "Translation and translation verification", 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.

Zhang, Y., T.N. Postlethwaite and A. Grisay (eds.) (2008), *A View Inside Primary Schools: A World Education Indicators (WEI) Cross-National Study*, UNESCO Institute for Statistics, Montreal.

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