Chapter 6

Analysis, reporting and use of data from international large-scale assessments in education

This last chapter looks at how the data from the reviewed large-scale assessments are analysed, reported and used. In particular, the chapter examines analytical approaches used for reporting, reports and communication of results, and use of data and results. In the case of each of the reviewed assessments, the chapter highlights any lessons that may be relevant for PISA for Development (PISA-D).

This chapter examines how the data from the reviewed surveys are analysed, reported and used, and considers the extent to which there is evidence that the data inform better teaching and learning.

It is divided into three subsections:

- analytical approaches used for reporting

- reports and communication of results

- use of data and results.

Note that more detail about the analytical approaches in general can be found in earlier sections of this document.

Analytical approaches used for reporting

The call for tender for the international contractor for Strand A and Strand B of PISA-D states that the contractor for Strand A should propose a methodology for scaling cognitive data that ensures comparability to PISA scales. It also states that data products and analytical outputs should include verified national datasets, a verified and adjudicated international dataset, and country-specific analytical outputs. Some of these outputs may be defined by the OECD for all PISA-D countries, and some may be defined in consultation with national project managers from individual countries (OECD, 2014: 44).

In PISA, cognitive data are scaled using a one-parameter item response theory model. Contextual data are used to create simple indices and scaled indices. The scaled contextual indices are also created using a one-parameter item response theory model. The scaled cognitive data are used to develop described proficiency scales. The proficiency scales consist of numeric PISA scores divided into segments or "levels", and substantive descriptions of the skills and abilities that correspond to each level. That is, proficiency levels are not an indication of students' performance relative to one another. Each proficiency level describes what students performing at that level know and can do. The substantive descriptions are developed based on inspection of the content and process demands of the items.

The IEA studies prePIRLS, PIRLS and TIMSS also develop proficiency scales for their cognitive data. PIRLS and TIMSS report achievement results for participating countries and benchmarking entities overall, as well as for the separate processes (for PIRLS), or cognitive/content domains (for TIMSS). These surveys compare means and distributions of student performance on the relevant subdomains to performance on the domain overall. Both PIRLS and TIMSS analyse trends in performance overall and trends in performance of different genders. Because TIMSS tests at both Grade 4 and Grade 8, it is also able to conduct cohort comparisons; that is, TIMSS compares the fourth grade results of the previous cycle and the eighth grade results of the current cycle to examine cohort progress over time. (For examples of analysis used in reporting for PIRLS and TIMSS, see Martin et al., 2012; Mullis et al., 2012a; Mullis et al., 2012b.)

One key aspect of the analysis and reporting of performance data from these surveys is the use of benchmarks. The benchmarks are unchanging points along the achievement scale: the advanced international benchmark (at 625 points on the scale), the high international benchmark (at 550 points), the intermediate international benchmark (at 475 points) and the low international benchmark (at 400 points). A scale anchoring exercise is employed each assessment cycle to describe student competencies at each of the

benchmarks (Mullis, 2012). Benchmarks are not an indication of students' performance relative to one another; it is theoretically possible for all students to reach or exceed a given benchmark. The benchmarks are linked to proficiency levels: for example, the high international benchmark maps to proficiency levels for each domain, and the levels describe what students generally must know and be able to do to be considered to be performing at a high standard. For each country, analyses are conducted to report on percentages of students reaching each international benchmark and trends in percentages of children reaching them.

When it comes to linking performance and contextual factors, the IEA studies explore contextual factors in themes that draw on data from the different questionnaires. In addition, for the first time in 2011, TIMSS and PIRLS developed policy-relevant scales covering areas including resources available at home for learning and education, resources available at school, teacher working conditions, school climate and students' attitudes towards learning. The surveys compared locations on these policy-relevant scales to performance (Mullis et al., 2012c).

The prePIRLS and PIRLS 2011 national level report for South Africa may be of interest to PISA-D (Howie et al., 2012). The analyses in this report are restricted to computing mean performance and comparing differences in mean performance over time and between groups divided by contextual variables of interest. No correlations, variance analyses, regression analyses or multivariate analyses are conducted.

In SACMEQ, the performance of students and teachers is analysed using Rasch model item response theory. In one working paper that presents international results for SACMEQ III, mean scaled scores are calculated and compared across countries, genders, and socio-economic status derived from a scaled SES indicator. Scores are also reported in relation to the eight competency levels that have been identified for each of the domains of reading and mathematics (Hungi et al., 2010). In another working paper that presents international results, multi-level analyses have been conducted to identify key student and school-level factors that influence achievement, to explore within-school and between-school variations in achievement, and to examine how social and gender differences in achievement compare after controlling for other factors that influence achievement (Hungi, 2011a). Other working papers from SACMEQ III calculate frequencies for contextual variables of interest, but do not connect these contextual factors to performance (see, for example, Hungi, 2011b; Hungi, et al., 2011). In the national-level SACMEQ reports, mean achievement scores are calculated for groups divided by contextual variables of interest such as region, sex, school location (rural/urban), and socio-economic status (low/high) (for example, see Monyaku and Mmereki, 2011). Changes in mean scaled scores across multiple implementations of SACMEQ are also calculated, as are percentages of scores falling within the eight competency levels defined for each of the two domains.

LLECE reports assessment results using a single continuous scale for each domain obtained from the application of the Rasch model item response theory. LLECE uses hierarchical linear modelling to analyse factors associated with student achievement, in order to contextualise results. Hierarchical linear modelling is a complex form of ordinary least squares (OLS) regression that is used to analyse variance in the outcome variables when the predictor variables are at varying hierarchical levels; for example, students in a classroom share variance according to their common teacher and common classroom

LLECE's strategy for analysis and reporting consists of two stages. In the first stage, LLECE publishes a report with the overall results for the region and each country,

focusing on comparing the average scores of countries and variance in each of the assessed grades and subjects.

In this first stage, results are also analysed in terms of performance levels describing what students can do. LLECE has four performance levels for each grade. These levels are specified simultaneously for each content domain and cognitive process that is assessed, and reflect progressive levels of difficulty. Countries are compared based on the percentage of students reaching each of these levels (LLECE, 2008).

In the second stage, normally two or three years after the assessment has been completed, LLECE publishes a report on associated factors, aiming to explore the relationship between student and school variables (obtained from the context questionnaires) and student achievement (for example, see Treviño et al., 2010). The purpose of the second stage of analysis and reporting is not only to relate contextual factors to student performance, but also to identify influential factors that could be modified by educational policy, particularly at the school level.

WEI-SPS uses several analytical approaches to survey data. These include: calculating proportions; correlations between variables and indices; analysis of variance by categories such as region and school type; mean values of indices; differences between mean values reported as effect sizes; and factor analyses. Even though surveys were administered to school leaders and teachers only, student weights were applied to leader and teacher data in order to report at the student level (Zhang, Postlethwaite and Grisay, 2008).

For EGRA and EGMA, implementing organisations choose the analyses that best serve their purposes. The guidance notes on planning and implementing EGRA provide some examples of the kinds of analysis and reporting options that implementing countries might choose (RTI International and International Rescue Committee, 2011).

- The letters, words and non-words subtasks and the phonemic awareness subtask analyse mean score per minute, disaggregated by groups (such as grade, gender and region).

- The listening comprehension subtask analyses the percentage of questions answered correctly disaggregated by groups.

- The oral reading fluency with comprehension subtask analyses: mean number of connected text words per minute disaggregated by groups; percentage of zero scores (a zero score is a child who can read no words in the oral reading passage correctly within the allocated time for the task); average percentage of correct answers; and percentage of children reading with at least 80% comprehension (a Fast Track Initiative indicator).

To give an example of the analytical options chosen for specific country implementation of EGMA, the baseline report from Kenya's Primary Math and Reading Project has the following (RTI International, 2012):

- The number identification subtask analyses the mean number of correct numbers identified disaggregated by cohort, location, sex and grade.

- The number discrimination subtask and missing number/pattern subtask analyse the mean percentage correct, disaggregated by cohort, location, sex and grade.

- Addition levels 1 and 2 analyse mean scores disaggregated by cohort, location, sex, grade, and percentage of students with zero scores.

- Subtraction levels 1 and 2 analyse mean scores disaggregated by cohort, location, sex, and grade, and percentage of students with zero scores.

- Word problems analyse the percentage correct, disaggregated by cohort, location, sex and grade.

Regarding the household-based surveys, PIAAC used item response theory to create proficiency scales and then defined proficiency levels for those scales. Six proficiency levels were defined for literacy and numeracy (Levels 1 through 5, plus below Level 1) and four for problem-solving in technology-rich environments (Levels 1 through 3, plus below Level 1). Differences in proficiency were compared across countries and across key socio-demographic factors (such as gender, educational qualifications, socio-economic status and occupation). Proficiency was also explored with reference to wages, labour market status and social outcomes (OECD, 2013).

In STEP, data presented in the reports include pass/fail information for the core assessment, the reading component score(s) and timing data, and information on the target population's reading literacy level, which is provided on the same five-level scale as used in the PIAAC literacy assessment (Pierre et al., 2014: 44).

In LAMP, performance data are analysed using item response theory to obtain proficiency scales. Three described proficiency levels were developed for the three domains (the prose domain, the document domain and the numeracy domain). In the national reports of LAMP results (for example, see UIS et al., 2013), frequency analyses are conducted to report on percentages in each proficiency level on overall data and on data disaggregated by gender, age and level of education.

In ASER, data are not analysed using item response theory. Each child receives a score that indicates the highest level he or she attained in the reading or maths test. These scores constitute the performance data. Performance is then analysed via frequency analyses to report the percentage of children attaining each level in the reading and mathematics assessments. Performance frequency analyses are conducted for reading and maths separately on data disaggregated by age and school type (see, for example, ASER Centre, 2014). The lowest level at which results are reported is the district. Performance trend analyses examine changes in the percentages of children in different grades attaining particular performance levels over time. The 2013 annual report also described changes in percentages of children attaining particular performance levels by tuition status. Little is done to link performance to family background, and no attempt is made to connect the results on the reading and mathematics assessments to any school characteristics except school type.

In Uwezo, as in ASER, data are not analysed using item response theory, and each child receives a score indicating the highest level he or she attained in the test. At the regional level (see, for example, Hoogeveen and Andrew, 2011; Uwezo, 2012a, 2014), analysis consists of comparing "pass" rates between countries and districts. A child is said to have passed the test if he or she got all attempted tasks correct. Test pass rates are also compared across socio-economic groups, with the index of socio-economic status calculated as a simple sum of responses to questions about durable assets owned, access to electricity and clean water, and mother's level of education (Uwezo, 2012a: 18). Trend analysis consists of comparing pass rates and frequency of children completing each task over time. At the national level (see, for example, Uwezo Kenya, 2011, 2013a; Uwezo Tanzania, 2010, 2011, 2013; Uwezo Uganda, 2010, 2011, 2013a), analyses focus on differences in performance between districts and regions within the country, in terms of

the percentage of students successfully completing each task. The relation between contextual variables and performance is not consistently investigated across countries or across years.

PASEC uses proficiency scales for cognitive data. Results are reported for participating countries and benchmarking entities overall and for the separate processes or cognitive and content domains. Means and distributions of student performance on the relevant subdomains are compared to performance on the domain overall. For each country, analyses are conducted to report on percentages of students reaching each international benchmark and trends in percentages of children reaching each international benchmark. Performance on specific items is explored. Score differences are presented with reference to schools, teachers and students factors. Time benchmarking will be proposed across PASEC's cycles.

Reports and communication of results

The call for tender for the international contractor for Strand A and Strand B of PISA-D discusses a technical report that will be prepared by the contractor, an international report that summarises overall international results that will be prepared by the OECD Secretariat, and national-level results reports. Regarding the preparation of national reports, the international contractor for Strand A is expected to provide analytical outputs, provide feedback and suggestions about analytical outputs, support the OECD Secretariat and participating countries during the development of the national reports, and review and provide comment on the draft national reports. The call for tender refers to the OECD PISA series *Strong Performers and Successful Reformers in Education* as giving examples of the types of analysis and reporting approaches that might be used for PISA-D national reports. (OECD, 2014).

For PISA, the OECD also prepares policy-oriented notes in a series called *PISA in Focus* (OECD, 2015). Shorter documents of the same kind are not mentioned in the call for tender international contractors for Strand A and Strand B of PISA-D.

Regarding the reviewed large-scale international surveys, the IEA studies prePIRLS, PIRLS and TIMSS communicate results through international reports prepared by the TIMSS and PIRLS International Study Center. The study centre also compiles the TIMSS and PIRLS encyclopaedias. The encyclopaedias present data from the curriculum questionnaires answered by national centre representatives from participating countries. These data are not analysed, but simply presented in a way that enables easy comparison. In addition, for the encyclopaedias, each country prepares a chapter summarising the structure of its education system, the language and reading curriculum in the primary grades, and overall policies related to reading instruction (such as teacher education, instructional materials and assessment) (Mullis and Martin, 2013: 6). IEA also produces technical reports that describe in detail all technical aspects of its assessments. These technical reports are published online.

The TIMSS results, reports, encyclopaedias, technical reports, assessment frameworks and other documentation for all cycles can be downloaded from the website of the TIMSS and PIRLS International Study Center.[1] The international databases for all cycles, and accompanying user guides, can be downloaded from the TIMSS and PIRLS website.[2]

IEA's Data Processing Center has developed the IEA International Database Analyser and IEA Data Visualiser software applications to facilitate the analysis and visualisation of data from IEA studies. These applications can be downloaded from IEA's website.[3]

Participating countries publish national reports to disseminate findings to a wide range of audiences within those countries, including government officials, policymakers, researchers and educators. These reports present national results in an international context, highlighting issues of special interest in the specific education system.

In SACMEQ, the students' reading and mathematics achievement scores for each of the three SACMEQ studies (I, II and III) are shown by country on the SACMEQ's website. Mean scores and standard errors for each of the subjects are disaggregated by region and other subgroups (gender, school location and SES).[4] All types of reports and data files from the three SACMEQ projects are also publicly available on SACMEQ's website.

The SACMEQ Coordinating Centre releases a number of international working papers with cross-national comparison and descriptions of technical aspects of the SACMEQ studies. For example, the topics of the SACMEQ III study working papers include: pupil achievement levels in reading and mathematics; performance levels and trends in school resources among SACMEQ school systems; characteristics of Grade 6 pupils, their homes and learning environments; characteristics of Grade 6 teachers; characteristics of school heads and their schools; trends in the magnitude and direction of gender differences in learning outcomes; and accounting for variations in the quality of primary school education.

Each SACMEQ participating country issues a policy brief and a detailed country report. Within the country report, sections are devoted to describing the background of the education system, the administration of the study, contextual information, the performance of students and teachers, and policy recommendations (SACMEQ, n.d.). It appears that the SACMEQ Coordinating Centre provides participating countries with considerable assistance in writing national reports. For example, a source version of a chapter for inclusion in a SACMEQ national report is available for download from SACMEQ's website. It is written by members of the SACMEQ Scientific Committee and staff from the International Institute for Educational Planning. Most of the participating countries seem to have used the source version in their national reports with few adaptations. Moreover, the three SACMEQ III workshops in 2009, 2010 and 2011, were all devoted to some aspect of national report preparation.[5] PISA-D participating countries might require similar levels of support in national report preparation.

When it comes to dissemination activities beyond the main reports, each SACMEQ country convenes research results dissemination forums for different groups of stakeholders, ranging from high-level policymakers and senior management of education ministry to donor agencies and regional and local level decision-makers (Nzomo and Makuwa, 2006).

LLECE's strategy for reporting results consists of two stages – an overall report is produced in the first stage and a report that explores contextual factors associated with performance in more depth is produced in the second stage. An important aim of the second-stage report is to identify factors that are both influential and might be modified by changes in educational policy.

WEI-SPS has produced a report that presents overall findings by theme, and then presents country profiles at the end of each theme section (see Zhang, Postlethwaite and

Grisay, 2008). There is also a WEI-SPS technical report (see UIS, 2009a). WEI-SPS data have been incorporated into the database compilation available on the UIS website.[6]

With respect to the other school-based surveys, EGRA and EGMA dissemination is done via the EdData website as well as, in the majority of cases, in-country seminars and discussions with key stakeholders.[7] Implementing organisations are expected to share their reports and instruments with RTI so they can be posted on this website.

In EGRA and EGMA, reporting and communication varies from implementation to implementation. The guidelines for planning and implementing EGRA discuss the different potential audiences at different levels (international, national, regional, community, school) and how dissemination activities can be targeted to these different audiences. The example dissemination products and activities the guidelines describe are: policy dialogue workshop; policy brief; social mobilisation campaign; project revision meeting; events with schools or communities; and teacher professional development (see RTI International and International Rescue Committee, 2011: 78-87).

Regarding the reviewed household-based surveys, PIAAC has a range of materials that are made publicly available on the OECD website.[8] In addition to a first results report that summarises international findings, there are interactive datasets and country notes, and links to national reports if they have been prepared.

STEP data, technical documents and national-level reports are available for download from the World Bank's data website.[9] The brochure for STEP states that national and international technical seminars will be organised to discuss the findings with national experts, including government officials, leading academic scholars, industry leaders, labour representatives and development partners (World Bank, 2012).

LAMP materials prepared by UIS suggest that an international database is available for download on the website.[10] The draft international planning report states that after analysis the UIS and the national teams must together undertake to develop statistical products and services that address the needs of different stakeholders, and formulate dissemination and communication strategies (UIS, 2004: 43-44). It also states that national teams will be expected to produce at least a national results report, a national technical report and a national micro dataset.

ASER prepares an annual results report and press statements. The release is televised. The ASER Centre website provides a lot of information related to the survey.[11] Some examples of what is available include: sample assessment tools; information about sampling; technical papers related to the survey; descriptions of sampling and the steps taken to ensure data quality; annual reports; tables of state- and district-level estimates; text from and links to articles discussing the survey from newspapers, magazines and online publications; information about the way ASER has featured in government policy and planning documents; lists and links to external publications that have made use of data from the survey; and a data query facility that presents state-level summary enrolment data and performance results for each year.

As mentioned above, Uwezo prepares regional and national reports. These materials are available for download from the Uwezo website.[12] The regional reports include frequency tables and graphs and charts comparing pass rates per country, divided by domain (English, numeracy and Kiswahili), age group and socio-economic level, and comparing trends over time. There are also district ranking tables that give districts with the highest pass rates and districts with the lowest pass rates, and an overall ranking table listing all districts in the three countries according their pass rates. There is no separate

technical report for Uwezo. The regional report includes some technical details, but the real emphasis is the main findings.

In the Uwezo national reports, frequency tables and charts giving the percentage of children able to complete each task are presented. Results are disaggregated by grade level, gender, age group and districts or regions. The emphasis of the reports depends on within-country decisions. The national reports include more explanation on the characteristics of the survey (such as sampling and test administration), but the level of detail varies across the three countries.

Uwezo has a well-articulated strategy for communicating results (Uwezo, 2012b). Of note is that the first step occurs during test administration; when test administrators give feedback to children and parents immediately after the test has been administered. At this point, test administrators also supply materials with practical steps that can be taken to improve learning (Uwezo Kenya, 2013b; Uwezo Uganda, 2013b; Uwezo, 2011, 2013a, 2013b).

In PASEC, the PASEC Centre is in charge of producing international reports (one per grade) while countries are in charge of their national report. Scores are the responsibility of PASEC Centre. International reports focus on international comparison and factors analysis. Technical documentations, framework and procedures are produced by PASEC Centre. While overall results are provided to meetings of PASEC and CONFEMEN, the PASEC Centre provides each country with its own database and scores, and each country prepares its own national report.

Among the reviewed surveys, there are efforts to explore innovative approaches to raising awareness and disseminating results to different constituencies.

The ASER Centre has its own capacity-building unit (ASER, 2015) that offers courses in basic descriptive statistics, Stata software and monitoring and research design. While these courses are intended to be general in nature, the ASER survey acts as the lens through which they are delivered. In that sense, they serve as a means of familiarising more researchers with the survey and its results.

Uwezo has identified radio as its preferred medium for reaching out to teachers and parents, and briefings have featured on a number of key radio stations in the region in which the survey is conducted.

IEA holds research conferences every few years (IEA, 2015) and a number of presentations at these conferences deal with PIRLS and TIMSS data.

In 2013, PASEC established a network of policymakers and technicians from the CONFEMEN countries. One aim of this network is to promote the use of assessment data in planning in the education sector. This network held a meeting for policymakers in 2014 (for a summary, see CONFEMEN, 2014).

In SERCE, a series of documents for teachers was prepared called '*Aportes para la enseñanza*', 'Contributions to teaching'. The series included one document for each domain that was assessed: reading, mathematics, writing and sciences. The aim of this series was to provide teachers with guidelines to improve their teaching strategies in the domains.[13]

UNESCO Santiago has a YouTube channel that hosts videos about the results from TERCE.[14]

Use of data and results

Given the large number of participating countries and the international renown of the OECD, PISA is arguably one of the most publicised and influential assessments in the world.

Representatives of participant countries often indicate that PISA has been used as a point of reference to modify national curricula or the focus of national assessment systems (Breakspear, 2012; Hopkins et al., 2008). In a number of countries, PISA results have even served to justify the introduction of or legitimise mass standardised testing procedures – not just for students but also for teachers – promoting the use of test results for public accountability (Froese-Germain, 2010).

In particular countries, PISA has influenced public policy differently. In countries that perform well (Finland and New Zealand, for example), PISA has not received excessive media coverage – governments have used results mainly as an external legitimisation for the organisation of their education systems or for justifying recent or upcoming educational reforms (Froese-Germain, 2010; Grek, 2009; Martens et al., 2010).

In countries performing below the national expectations, PISA results have caught the interest of public opinion. Two well-known cases of countries that have experienced what is known as the 'PISA shock' are Germany and Norway (Grek, 2009; Hopfenbeck et al., 2013; Martens et al., 2010).

The most documented cases about the use of PISA data to inform educational policy are from high-income countries with technical capacity and resources to handle project implementation and ensure that results feed into policymaking. Not all PISA countries have such capacity or resources. The OECD's working paper on PISA in low- and middle-income countries discusses two participating countries that do not fit into this category – Tunisia and Kyrgyzstan. The paper quotes Tunisia as giving "lack of political will and know-how" as the reason why better use is not made of PISA national data. In Kyrgyzstan, limited local capacity and a sense that the national education has been shaped by external interests mean that PISA and its results are largely unknown by the general public and national education stakeholders (Bloem, 2013).

Regarding the reviewed large-scale international surveys, countries use PIRLS and TIMSS achievement data for system-level monitoring in a global context, and monitoring progress in achievement over time (Mullis et al., 2012d, 2013).

The PIRLS 2011 encyclopaedia states that countries with low performance compared to other countries have initiated educational reforms in response to the results, and that countries with declining performance have sometimes formulated new goals and policies to drive improvement. The PIRLS 2011 encyclopaedia also highlights that the surveys encourage many countries to make special efforts to address any equity issues that are revealed by the results, and that the surveys often motivate countries to improve classroom instruction as well (Mullis et al., 2012d: 17-18). The TIMSS encyclopaedia describes a similar situation (Mullis et al., 2013: 25-26).

South Africa's response to the PIRLS 2006 results may be of interest to PISA-D. A cluster of initiatives appears to have been influenced by the PIRLS 2006 results, based on the time at which they were implemented. Initiatives ranged from increased library funding, to the development of a national reading strategy, to handbooks for teachers. The PIRLS data from 2006 function as a baseline level against which the success of these initiatives can be monitored (Howie et al., 2012: 15-16).

SACMEQ research results have been playing an important role in informing dialogue and decisions related to the education systems of the member countries (Leste, 2005; Nzomo and Makuwa, 2006; Sayed and Kanjee, 2013). When SACMEQ I was completed, for example, the project reports featured in major policy documentation such as: presidential and national commissions on education in Kenya, Namibia and Zimbabwe; a prime ministerial and cabinet review of educational policy in Zanzibar; national education sector studies in Malawi and Zambia; and a review of a national education master plan in Mauritius (Murimba, 2002).

Moreover, the influence of SACMEQ research results can be observed not only in policy documentation, but also in the actual direction of policy and practice reforms in some countries. In Kenya, for example, SACMEQ findings on lower-than-expected levels of achievement have prompted the government, in collaboration with other key stakeholders and development partners, to implement a school-based teacher development programme. Donors have also begun to support the provision of textbooks to all public primary schools when findings showed there was an inadequate supply of them (Nzomo and Makuwa, 2006).

Another example is Namibia, where findings from the SACMEQ research revealed that the northern regions had the most difficulty in providing adequate educational resources and achieving minimum levels of student learning outcomes. With the support of development partners, multiple levels of the education sectors in these regions – from teachers to regional education officers – have now been targeted for assistance. Schools have been divided into clusters for administrative and support services. This arrangement enables a cluster of schools to share educational resources, good practice, and valuable expertise, which can benefit struggling schools in the region (Nzomo and Makuwa, 2006).

In both examples from SACMEQ, active involvement by ministry of education staff in the research implementation was key in linking results and action (Nzomo and Makuwa, 2006).

It is difficult to determine the extent to which LLECE results and data have influenced efforts to improve teaching and learning in participating countries. Though LLECE reports always conclude with a chapter on recommendations for education policy development (see, for example, LLECE, 2013), no information is available about whether these recommendations have triggered any changes in policy or practice. A study on this topic is envisaged in the LLECE Strategic Plan for 2015–2019 (M. Bilagher, personal communication, February 2014). LLECE is now developing a methodology to use study data for policy development at the micro and at the macro level.

Results from WEI-SPS provide a descriptive portrait of reported teaching practices in fourth grade literacy and mathematics in the participating countries, but since no cognitive assessments were administered, these practices cannot be evaluated against learning outcomes. Knowing what is occurring in their own and other countries can better help education reform stakeholders (policymakers, jurisdictional authorities and educators) identify possible gaps between planning and programme formulation and actual service delivery.

With respect to the other school-based surveys, EGRA and EGMA are often used to evaluate the impact of an initiative (baseline and endline studies are undertaken). In these instances the results can be said to inform teaching and learning because they can help to

show that some practice leads to improved outcomes or show that the practice makes no difference.

There are other examples of how EGRA data have been used to inform teaching and learning. In 2008, Nicaragua undertook a national level diagnostic assessment of reading using EGRA. The aim was to analyse the reading ability outcomes of children in the early grades and to examine the contextual factors that may be responsible for the observed outcomes. After the EGRA results were analysed, Nicaragua's ministry took immediate, positive steps to address the quality of instruction, and also refocused its attention and efforts on quality improvements in the early grades (Gove and Wetterberg, 2011).

Additionally, in 2009, Liberia used EGRA as the primary source of data to inform instruction and to gauge efficacy of reading instruction at the individual, classroom, school, family and community levels. A modified, curriculum-specific EGRA was used as a classroom tool for continuous assessment. This classroom assessment tool facilitated setting reading performance goals and provided a benchmark for teachers, schools, administrators, families and other community stakeholders could use to evaluate classroom reading instruction. EGRA tools also provided a link to instruction as teachers could assume that students' scores on the EGRA measures were directly related to the general reading outcome goals, and that increased scores meant that the reading instruction contributed to students' learning. If there was no increase in student scores over time, then teachers understood that they needed to modify instruction (Gove and Wetterberg, 2011).

Regarding the reviewed household-based assessments, the STEP results have not been released for long, so there has most likely not been enough time for evidence to be produced about how they are influencing efforts to improve teaching and learning. The results are being incorporated into World Bank reports and policy discussions.

The guidelines for implementing LAMP discuss some ways it is envisaged that the survey data will feed into efforts to raise literacy levels. In particular, the survey data will support the design of literacy programmes and the improvement in educational policies by identifying the skills of the population (UIS, 2009b: 41).

In ASER, the ASER Centre website addresses the impact of the survey with a page that presents an archive of all state, national and international media coverage the survey has received. This page contains references that can be viewed by date or type of coverage. This widespread attention would undoubtedly raise awareness in the general public about children's learning levels. The website also lists when the findings of the survey and the issue of learning outcomes in general have featured in education policy and planning discussions, and refers to how many district level teacher training institutes supply volunteer field investigators for the survey. The website also has a page that gives brief information about the assessments introduced in other countries that are based on ASER principles and methodology. The ASER Centre has played an important role in supporting these initiatives.

In addition to the material available on the ASER Centre website, our attention has been drawn to an initiative in Jehanabad district in Bihar. In this initiative, the ASER tools were used to determine children's reading levels and to support teachers in targeting teaching to these reading levels rather than above them, as the curriculum might dictate. This approach was subsequently scaled up to cover other areas in the state of Bihar (R. Banerji, personal communication, 19 June 2014).

There is little information available about how the Uwezo approach and results are used in efforts to improve teaching and learning. There are, however, two instances of interest. In one instance, the Ikhoba Girls Primary School in Masindi district in Kenya redoubled their efforts to improve learning levels after poor district level results in Uwezo 2011. In another instance, a district co-ordinator in Homa Bay Town in Kenya used connections established through Uwezo to form village education committees to facilitate better communication between families, school management staff and teachers about local education issues (S. Ruto, personal communication, 31 August 2014).

The impact of Uwezo is considered in a study undertaken in two rural Kenyan districts by researchers at Princeton University and Massachusetts Institute of Technology. The study found that two of the most valued Uwezo strategies – the instant feedback provided to parents regarding their children's competencies, and the provision of support materials to improve learning – had no impact on increasing citizen activism. Based on a series of measures of citizen activism, the study concluded that parents who received the instant feedback and the support materials were not more likely to act to improve the quality of their children's schooling or to adopt behaviours at home that might improve learning than parents who did not receive the information (Lieberman, Posner and Tsai, 2013). Although discouraging, the researchers suggest that their findings may indicate that Uwezo's provision of information is failing to trigger behavioural change due to other more general factors influencing the causal relationship between information provision and action. These factors may include the ability to understand the information that is being received, the level of responsibility that the audience feels regarding the information, and the level of belief people have that their actions will generate results.

Implications

Analytical approaches used for reporting

We suggest that the OECD examines the use of benchmarks in the reviewed surveys and considers whether benchmarks might be incorporated into PISA-D analysis and reporting. Benchmarks that define minimum expected levels of performance may become increasingly relevant in the context of the post-2015 development goals and targets for education quality.

We also suggest that the OECD makes sure that questionnaire scales developed and used in PISA-D reporting are considered relevant to policy in the participating countries.

Additionally, we suggest that the OECD refers to national level reports from relevant countries that have participated in the reviewed large-scale assessments (such as South Africa in prePIRLS and PIRLS 2011, the SACMEQ countries). These reports provide a sense of the kinds of analysis and reporting options that these countries have deemed relevant for their contexts, and that may be relevant for PISA-D.

Reports and communicating results

We suggest that the OECD considers whether a presentation of participating country contexts such as that given by the TIMSS and PIRLS encyclopaedias may be valuable for PISA-D. Other surveys or monitoring efforts may already have systems in place to capture at least some of this information.

We suggest that the OECD and the international contractors for Strand A and Strand B of PISA-D should be prepared to offer considerable support to countries for the important work of preparing national results reports.

We also suggest that the OECD considers supporting participating countries to develop and implement dissemination plans. In many of the reviewed assessments, there was very little national level material available. Without national level material that is judged by decision makers as useful and relevant, a survey can only ever have a limited impact.

Use of data and results

We suggest that the OECD takes note of the observation from SACMEQ that active involvement of ministry staff in the research implementation is the key to linking results and actions, and considers how to ensure that government buy-in leads to similar success with PISA-D.

Notes

1.   See http://timssandpirls.bc.edu/isc/publications.html.

2.   See http://timss.bc.edu.

3.   See www.iea.nl/data.html.

4.   See www.sacmeq.org/sacmeq-projects/sacmeq-iii/readingmathscores.

5.   See www.sacmeq.org/training-workshops – one workshop was about accessing and analysing data files for national reporting, another was about preparing draft chapters for national reports, and another was about sharing, reviewing and improving draft chapters for national reports.

6.   See http://data.uis.unesco.org.

7.   See www.eddataglobal.org/index.cfm.

8.   See www.oecd.org/site/piaac.

9.   See http://microdata.worldbank.org/index.php/catalog/step.

10.   See www.uis.unesco.org.

11.   See www.asercentre.org/ - 6dwi6.

12.   See www.uwezo.net.

13.   See www.unesco.org/new/es/santiago/education/education-assessment-llece/second-regional-comparative-and-explanatory-study-serce.
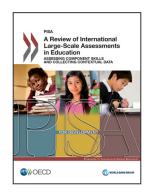
14.   See www.youtube.com/user/UNESCOSantiago.

References

ASER (2015), "ASER Centre capacity building", https://sites.google.com/site/asercentrec apacitybuilding (accessed 23 May 2015).

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

Bloem, S. (2013), "PISA in low and middle income countries", *OECD Education Working Papers* No. 93, OECD, Paris, http://dx.doi.org/10.1787/5k41tm2gx2vd-en.

Breakspear, S. (2012), "The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance", *OECD Education Working Papers* No. 71, http://dx.doi.org/10.1787/5k9fdfqffr28-en.

CONFEMEN (2014), "Subregional workshop on the role and place of assessment in education systems' steering and reform: Policymakers workshop", summary report, CONFEMEN, Dakar, Senegal.

Froese-Germain, B. (2010), *The OECD, PISA and the Impacts on Educational Policy*, Canadian Teacher's Federation, Ottawa.

Gove, A. and A. Wetterberg (eds.) (2011), *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy*, RTI International, North Carolina.

Grek, S. (2009), "Governing by numbers: The PISA 'effect' in Europe", *Journal of Education Policy* 24(1), Routledge, London, pp. 23-37.

Hoogeveen, H. and D. Andrew (2011), *Are Our Children Learning? Numeracy and Literacy across East Africa 2011*, Uwezo, Nairobi.

Hopfenbeck, T. et al. (2013), "Balancing trust and accountability? The Assessment for Learning Programme in Norway: A governing complex education systems case study", *OECD Education Working Papers,* No. 97, http://dx.doi.org/10.1787/5k3txnpq lsnn-en.

Hopkins, D. et al. (2008), *The Global Evaluation of the Policy Impact of PISA*, OECD, Paris.

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.

Hungi, N. et al. (2011), "SACMEQ III project results: Levels and trends in school resources among SACMEQ school systems", *SACMEQ Working Document*, SACMEQ, Paris.

Hungi, N. et al. (2010), *SACMEQ III Project Results: Pupil Achievement Levels in Reading and Mathematic*s, SACMEQ, Paris.

IEA (2015), "IEA International Research Conference", www.iea.nl/irc.html (accessed 23 May 2015).

Leste, A. (2005), "Streaming in Seychelles: From SACMEQ research to policy reform", paper presented at International Invitational Educational Policy Research Conference, SACMEQ, Paris.

Lieberman, E., D. Posner and L. Tsai (2013), "Does information lead to more active citizenship? Evidence from an education information intervention in rural Kenya", *MIT Political Science Department Research Paper,* No. 2013-2*,* http://ssrn.com/abstract=2228900.

LLECE (2013), *Diseño Muestral Tercer Estudio Regional Comparativo y Explicativo (TERCE)* (Sampling Framework for the Third Regional Comparative and Explanatory Study (TERCE), LLECE, Santiago.

LLECE (2008), *Segundo Estudio Regional Comparativo de los Aprendizajes de los Estudiantes de América Latina y El Caribe. Primer Reporte* (Second International Comparative Study of Student Learning in Latin American and the Caribbean. First Report), Latin American Laboratory for Assessment of the Quality of Education (LLECE) / UNESCO-Santiago, Santiago.

Martens, K. et al. (2010), *Transformation of Education Policy*, Palgrave, Basingstoke, New York.

Martin, M. O. et al. (2012), *TIMSS 2011 International Results in Science*, TIMSS and PIRLS International Study Center, Chestnut Hill, MA.

Monyaku, B. and O.A. Mmereki (2011), *The SACMEQ III Project in Botswana: A Study of the Conditions of Schooling and the Quality of Education*, Botswana Ministry of Education and Skills Development, Division of Planning Statistics and Research, Gabarone.

Mullis, I.V.S. (2012), "Using scale anchoring to interpret the TIMSS and PIRLS 2011 achievement 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.

Mullis, I.V.S. et al. (eds.) (2013), *TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science, Volume 1: A–K*, TIMSS and PIRLS International Study Center, Chestnut Hill, MA.

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

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.

Mullis, I.V.S. et al. (2012c), "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. (eds.) (2012d), *PIRLS 2011 Encyclopedia: Education Policy and Curriculum in Reading, Volume 1: A–K*, TIMSS and PIRLS International Study Center, Chestnut Hill, MA.

Mullis, I.V.S. and M.O. Martin (eds.) (2013), *PIRLS 2016 Assessment Framework*, TIMSS and PIRLS International Study Center and IEA, Chestnut Hill, MA and Amsterdam.

Murimba, S. (2002), "SACMEQ education ministers review progress and future plans", *IIEP Newsletter*, Vol. XX, No.1, January-March, International Institute of Educational Planning, UNESCO, Paris.

Nzomo, J. and D. Makuwa (2006), "How can countries move from cross-national research results to dissemination, and then to policy reform? (Case studies from Kenya and Namibia)" in K. Ross and I.J. Genevois (eds.), *Cross-National Studies of the Quality of Education* IIEP, UNESCO, Paris, pp. 213-228.

OECD (2015), "PISA in Focus", www.oecd.org/pisa/pisaproducts/pisainfocus.htm (accessed 4 August 2015).

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

OECD (2013), *OECD skills outlook 2013: First results from the Survey of Adult Skills: OECD Publishing*, Paris, http://dx.doi.org/10.1787/9789264204256-en.

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

RTI International (2012), *The Primary Math and Reading (PRIMR) Initiative Baseline Report*, RTI International, North Carolina.

RTI International and International Rescue Committee (2011), *Guidance Notes for Planning and Implementing EGRA*, RTI International, North Carolina.

SACMEQ (n.d.), "SACMEQ projects" www.sacmeq.org/sacmeq-projects (accessed on 11 April 2014).

Sayed, Y. and A. Kanjee (2013), "Assessment in sub-Saharan Africa: Challenges and prospects", *Assessment in Education: Principles, Policy and Practice,* 20(4), Routledge, London, pp. 373-384.

Treviño, E. et al. (2010), *Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe* (Factors associated with student achievement in Latin America and the Caribbean), UNESCO-OREAL Santiago and LLECE, Santiago.

UIS et al. (2013), *LAMP: Country Summary for Paraguay,* UNESCO Institute for Statistics, Montreal, www.uis.unesco.org/literacy/Documents/LAMP%20Country%20 Summaries/literacy-statistics-summary-paraguay.pdf.

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 (2004), "Literacy Assessment and Monitoring Programme (LAMP): International planning report" (draft), UNESCO Institute for Statistics, Montreal.

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

Uwezo (2013a), "2012/2013 Preparation and training manual – Kenya", www.uwezo.net /assessment/training (accessed 10 March 2014).

Uwezo (2013b), "2013 Preparation and training manual - Uganda", www.uwezo.net /assessment/training (accessed 11 March 2014).

Uwezo (2012a), *Are our Children Learning? Literacy and Numeracy across East Africa 2012*, Twaweza, Nairobi.

Uwezo (2012b), *Standards Manual,* Uwezo, Nairobi.

Uwezo (2011), "Improving learning outcomes in East Africa 2009-2013: Strategy update", www.uwezo.net/strategies.

Uwezo Kenya (2013a), *Are Our Children Learning? Annual Learning Assessment Report 2012*, Uwezo and Women Educational Researchers of Kenya (WERK), Nairobi.

Uwezo Kenya (2013b), "Volunteer workbook - Kenya", www.uwezo.net/assessment /training (accessed 11 March 2014).

Uwezo Kenya (2011), *Are our Children Learning? Annual Learning Assessment Report 2011*, Uwezo, WERK, Nairobi.

Uwezo Tanzania (2013), *Are Our Children Learning? Annual Learning Assessment Report 2012*, Uwezo and Tanzania Education Network (TEN/MET), Dar es Salaam.

Uwezo Tanzania (2011), *Are Our Children Learning? Annual Learning Assessment Report 2011*, Uwezo, TEN/MET, Dar es Salaam.

Uwezo Tanzania (2010), *Are Our Children Learning? Annual Learning Assessment Report 2010*, Uwezo, TEN/MET, Dar es Salaam.

Uwezo Uganda (2013a), *Are Our Children Learning? Annual Learning Assessment Report 2012*, Uwezo Uganda, Kampala

Uwezo Uganda (2013b), "Volunteer workbook - Uganda", www.uwezo.net/assessment /training (accessed 11 March 2014).

Uwezo Uganda (2011), *Are Our Children Learning? Annual Learning Assessment Report 2011*, Uwezo and Uganda National NGO Forum, Kampala.

Uwezo Uganda (2010), *Are our Children Learning? Annual Learning Assessment Report 2010*, Uwezo and Uganda National NGO forum, Kampala.

World Bank (2012), "STEP Skills Measurement Study" (brochure), World Bank, Washington DC.

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.