HEART
HEALTH & EDUCATION ADVICE & RESOURCE TEAM

Helpdesk Report: Learning Assessments

QUERY:

What is the evidence on:

- how best to promote effective national capacities to conduct learning assessments?
- to what extent participation in international learning assessments has built national capacities to design, implement and make use of national assessments?
- participation in international learning assessments having an impact on political decisions, policy-making and teaching practices in countries?
- the consequences of focusing assessment of learning on language (reading), numeracy/maths and science?
- the circumstances and actions required to ensure learning assessments (both national and country participation in international assessments) promote and secure improvements in learning achievement?

Date: 1 October 2015

CONTENTS

List of Acronyms

EFA	Education for All
ICT	Information and Communication Technology
IEA	International Association for the Evaluation of Educational Achievement
INEADE	Institut National d'Étude et d'Action pour le Développement de l'Éducation
ILA	International Learning Assessments
LLECE	Latin American Laboratory for Assessment of Quality in Education
LMIC	Low- and Middle-Income Country
LSEA	Large-Scale Learning Assessment
NLA	National Learning Assessment
OECD	Organisation for Economic Co-operation and Development
PASEC	Programme for the Analysis of Educational Systems of CONFEMEN countries
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PISA-D	Programme for International Student Assessment for Development
RLA	Regional Learning Assessment
SABER	Systems Approach for Better Education Results
SACMEQ	Southern and Eastern African Consortium for the Monitoring of Education Quality
SDGs	Sustainable Development Goals
TIMSS	Trends in International Mathematics and Science Study
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UPE	Universal Primary Education

1. Introduction

This helpdesk report provides a rapid analysis of evidence of the role of large-scale learning assessments (LSEAs) in education systems in low- and middle-income countries. It is divided into five principal sections, each associated with one of the 5 sub-queries set out above. The information and analysis is supplemented by a number of Annexes detailing specific approaches to learning assessment design and implementation. A bibliography is included, with links for resources used.

The resources included in this report were identified through a non-systematic desk-based search. This report is a rapid response and, as such, it should be treated as a synthesis of the resources and evidence gathered in the assigned time.

2. Overview of key findings

Overall, the available evidence that publicly examines the links between large-scale assessments of students' learning and education policy, teaching practice and national capacity for assessment is limited to a small number of systematic reviews and a series of country-level case studies. Tobin et al. (2015) speculate that the reason for this small amount of published evidence might be that the bulk of evidence for such links is held in country- or programme-level briefings for governments, donors and other stakeholders, and is therefore either confidential or circulated at a programme level, and not easily available for public scrutiny.

Best et al. (2013) propose that future research and policy analysis may aim to examine the relationship between system-level factors as barriers and facilitators of education policy-making in developing countries, as, firstly, evidence of their impact is not available, and secondly, such factors are seen by the majority of commentators in this report as key in influencing the overall effectiveness and value of large-scale assessments. These factors included, for example, the effectiveness of the education system, political sensitivities and conflict, the strength of civil society and public discourse.

The query raised a number of important areas for research, which touch upon sophisticated themes that operate both at the macro-level of international assessment systems design and implementation, and also at the highly contextual micro-level of classroom-based assessment for learning. With this in mind, and as highlighted by the experts' comments, some topics covered here would benefit from further detailed analysis, and possibly, in the context of a rapid response report of this sort, being covered in isolation.

3. Background: International & regional learning assessments

International Learning Assessments

International Learning Assessments (ILAs) are primarily designed to focus on the measurement of learners in multiple countries. Their aims are to: develop cross-national comparisons that target a variety of educational policy issues; provide league tables that rank-order achievement scores by nation or region or other variables; measure international learning trends over time; monitor and evaluate the quality of student learning outcomes; and support within-country analyses that are then compared to how other countries operate at a sub-national level (Wagner 2011; Best et al. 2013)

Leading ILAs include the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA), and the International Association for the Evaluation of Educational Achievement's (IEA) Trends in International

Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). Wagner (2011) characterises these assessment programmes by their attention to high quality instruments, rigorous fieldwork methodology, and sophisticated analyses of results.

Of these ILAs, PISA is regarded as the principal source of international data on the performance and quality of education systems, as measured by student achievement outcomes (Froese-Germain 2010). PISA also has the highest visibility and participation rates among low- and middle-income countries (LMICs) (Best et al. 2013; Grisay et al. 2007). For example, 40 partner countries and economies participated in PISA 2009, a number greater than the total OECD countries (Bloem 2013). However, it is worth noting that, of these partner countries, the majority are grouped as upper middle-income. Further, to date, PISA participation from countries on the African continent has been limited, and there are no participant countries from sub-Saharan Africa (ibid).

In partial response to these circumstances, and in anticipation of the 2015 Sustainable Development Goals (SDGs), OECD developed PISA for Development (PISA-D), an ILA launched in 2013 and designed specifically to increase the participation of developing countries by improving its contextual relevance and policy focus – including on youth literacy, girls' education, and methodologies for assessment of out-of-school children (Bloem 2013).

Regional Learning Assessments

In addition to ILAs, there are a series of Regional Learning Assessment programmes (RLAs), undertaken to compare samples of schools in a region of the world in which countries may share similar economic and social conditions in order to explicitly compare student achievement cross-nationally (Best et al. 2013). The three major RLA programmes are: the Latin American Laboratory for Assessment of Quality in Education (LLECE); the Southern and Eastern African Consortium for the Monitoring of Education Quality (SACMEQ); and Program for the Analysis of Educational Systems of the CONFEMEN (Francophone Africa) countries (PASEC).

SACMEQ, covering East and Southern Africa, is seen as a high-profile example of an RLA (Best et al. 2013). Working across 15 countries since 1995, SACMEQ has conducted three sub-regional comparative studies of student performance in reading, mathematics and, in the most recent cycle, HIV and AIDS knowledge of sixth graders, based on common aspects of the 6th grade curriculum. It is currently developing a fourth cycle. The mandate of SACMEQ is to undertake research in order to generate information for policy making, with a strong focus on training activities to obtain the technical skills required to monitor, evaluate and compare empirical findings for educational planners and researchers within the ministries (Bloem 2013).

The development of the RLA programmes are seen as part of an effort to extend the use of large-scale education assessments (LSEAs) into developing countries (Wagner 2011). They function in similar ways to the ILAs, but they ensure that the regional dimensions are given substantive attention. These include: the relative proximity in content between the test and the regional curriculum; assessment scales that may be closer tied to localised skill levels; and greater attention to local policy concerns. These RLAs are given substantial credibility, and an increasing number of developing countries have participated in assessments undertaken at regional level (ibid).

4. Promoting effective national capacities to conduct learning assessments

Developing effective assessment systems is a 'complex and long-term undertaking that often requires high investments in human and financial resources' (Wagner 2011). According to Braun and Kanjee (2007) and Crooks (2002), countries with limited resources and technical expertise should undertake the development of effective systems in stages. Key issues of the assessment system, such as staff development, should be addressed, and in order to do this, policy makers need to carry out an analysis to assess capabilities and needs.

The promotion of effective national capacity to conduct learning assessment relies on the extent to which initiatives can identify and address issues associated with the above factors on a contextual basis. For example, in the case of SACMEQ, Murimba (2005) concludes that the three key factors that have influenced the magnitude of impact are: (a) the organizational culture of the host ministry; (b) opportunities created by other forces impacting on educational policy reform; and (c) technical and advocacy skills of the national co-ordinating bodies and their partners.

In terms of addressing these issues, Leitz and Tobin (2014) state that available evidence demonstrates that, the closer the link between an assessment program and national education policy makers, especially in the design phase, the greater the impact of assessment on educational policy making at the national level. More specifically, further practice-based research by Tobin et al. (2015) presents the following recommendations to promote effective national capacity for conducting learning assessment:

- Prioritise the integration of LSEAs within policy-making processes, for example:
    - Formally legislate the establishment and financing of assessment programs and agencies;
    - Ensure that data relevant to identified policy concerns is prioritised by the assessment;
    - Ensure that the assessment include questions about background factors related to student outcomes (e.g. socioeconomic background; availability of resources at school and home);
    - Facilitate regular meetings and seminars between assessment officials and policy-makers to facilitate communication, dissemination and understanding of results;
    - Ensure that the dissemination of assessment results includes targeted policy papers linked to priority concerns.
- Improve the technical quality of assessments, including developing the capacity of those involved in their design and implementation, for example:
    - Ensure that best practice is followed in assessment design and implementation;
    - Engage in ILAs or RLAs that emphasise national capacity building, in order that technical skills may be applied nationally;
    - Pursue capacity development opportunities for staff through regional networks, technical assistance agencies, university courses or other training programs.
- Ensure that assessments have a sound communication and dissemination strategy to engage stakeholders, including the media, for example:
    - Ensure the dissemination of assessment results to all stakeholders
    - Target dissemination according to the interests and technical knowledge of each stakeholder
    - Engage with the media through all phases of an assessment program in order to increase the media's understanding and facilitate better communication. (Tobin et al. 2015)

As a means of guiding and supporting such activities at both national and programme level, the World Bank has developed the SABER (Systems Approach for Better Education Results) initiative, a diagnostic toolkit which allows countries to measure their current capacities, plan

to further develop their assessment systems and learn from other policies and best practice from around the world. Its Framework for Building an Effective Student Assessment System identifies the key consideration of an effective system as 'the individual and combined quality of the assessment activities in terms of the adequacy of the information generated to support decision making' (Clarke, 2011).

Within SABER, the three key drivers of information quality in the national learning assessment are given as:

1. The Enabling Context
2. System Alignment
3. Assessment Quality

The Enabling Context refers to the broader context, including areas such as the policy/legal framework under which activities are carried out, the availability of funding and resources and the capacity and technical capabilities of staff. As the 'key driver' of quality and effectiveness, it is crucial to get the enabling context right (World Bank, 2010 cited by Clarke, 2011). It is usually up to the national government to provide the vision, leadership and framework of the enabling context, but governments can also collaborate on a regional level to jointly develop an effective enabling context (such as in SACMEQ).

System Alignment of an assessment programme into to the national education system is important to ensure that the assessment goals match the curriculum and teacher-training opportunities. Considerations may include i) domain coverage – to what extent assessments provide information on student learning and achievement ii) population coverage - how much of the student population is covered by the assessment activity and iii) utility – how useful and usable the assessment is for stakeholders (Clarke, 2011).

Assessment Quality refers to areas such as the design and implementation of activities and questions, the analysis and interpretation of student responses and the reporting of results. If quality is lacking in any of these areas then it may lead to poor decision-making (Wolff, 2007 in Clarke, 2011).

The diagnostic toolkit from the SABER initiative enables countries to position themselves in relation to the aforementioned areas as Latent, Emerging, Established or Advanced and provides generic examples (see adapted table below) to support countries with their self-positioning. Note that Latent is not included here as this supposes that little exists in the way of an assessment system.

	Emerging	Established	Advanced
Enabling context	• No or limited policy framework • Few trained staff • Unreliable funding	• Presence of policy framework • Training programs for staff • Stable/reliable funding	As with established, with a focus on: • Assessment for learning • School-based and classroom assessment • Role of teachers • Research and innovation
System alignment	• Assessments not fully aligned with learning goals, standards or	• Assessments aligned with learning goals, standards, curriculum	As with established, with a focus on: • Assessment for

	the curriculum • Limited use of results to inform policy and practice	• Systematic use of results to inform policy and practice	learning • School-based and classroom assessment • Role of teachers • Research and innovation
Assessment quality	• Limited awareness or application of technical or professional standards	• Awareness and application of technical or professional standards	As with established, with a focus on: • Assessment for learning • School-based and classroom assessment • Role of teachers • Research and innovation

According to Ferrer (2006), countries need to focus on reforms and practices that strengthen the Enabling Context for assessment, since enhancing institutional quality before assessment quality makes a learning assessment project more likely to succeed (Larach & Lockheed, 1992). For example, Chile developed its enabling context through strong political and technical leadership from the Ministry of Education and universities, political and economic stability, social pressures and curricular reforms (Ramirez, 2012 cited in Chulu, 2009). To move from Established to Advanced, it is recommended that a focus is put on reforms and practices that places classroom at the centre, and teachers and students as the key actors in assessment (Darling-Hammond and Wentworth, 2010; Shepard, 2000 in Clarke, 2011).

Case study: Institutionalising assessment capacity in Malawi

The Government of Malawi sought to institutionalise assessment capacity, by which it is meant 'routinising, in a systematic and coordinated manner, the assessment capacity-building programmes or practices in the country' (Chulu 2009). This was necessary because, between 1964-2004, there lacked an education policy regulating school-based assessment at the national level.

Employees at MANEB (the institution with responsibility for conducting national examinations) were trained at the University of Massachusetts (UMass) in advanced degrees (Masters and doctoral) and participated in workshops which took place in Malawi. Senior officers from MANEB also visited UMass and other organisations to learn from best practice in assessment. To increase sustainability, the postgraduate programme at Educational Measurement and Evaluation at Chancellor College in Malawi was reformed (to include modules in areas such as principles of educational measurement and fundamentals of test construction) and Malawi educators were trained in postgraduate degrees at UMass to become lecturers at Chancellor College (Chulu 2009).

Case Study: Developing the enabling context in Brazil
(taken from Guimaraes de Castro, 2012)

Over 20 years Brazil reformed its assessment system to develop one which enabled the reliable nationwide tracking of enrolment and learning levels. This in turn resulted in an improvement in the quality and efficiency of its system.

In order to enhance the enabling context, Brazil defined its policy framework for assessment, created an institution to lead assessment activities, committed sufficient resources to fund the programme and developed in-country capacity to undertake the activities. The factors that contributed to the success of reforms include stable political leadership, prioritising education quality and its monitoring, international support and partnering with nongovernmental institutions.

Key lessons for other countries that emerged from Brazil's case include:

- Developing an effective national assessment system is a long-term project that requires gradual implementation
- The importance of political leadership
- The need for a clear policy framework and ownership of the system by a designated body
- Ensuring that the strategic agency is not overloaded with implementation/administrative tasks
- Building staff capacity to run assessment programmes

More detailed information about this case study is available on the World Bank website.

5. Participation in ILAs to build national capacities for design, implementation and use of national assessments

The potential for capacity building is often mentioned as an important benefit of participation in ILAs, and examples frequently focus on the development of national technical and skills-based capacity related to development of assessment items, survey tools, sampling procedures, quality assurance, planning and implementation of data collection, and the secondary analysis of data sets. There is some evidence that participation by LMICs in international assessments may contribute to capacity building of this nature (Lockheed, 2010; Cariola et al., 2011 in Bloem 2013).

However, it is also noted that there are significant challenges to LMICs in taking full advantage of capacity-building opportunities that can feed into the development of country-specific National Learning Assessment programmes (NLAs) (Bloem 2013).

In general terms, these challenges include the following:

Costs of participation

As an example, the cost for participation in the PISA 2015 cycle was EUR 182,000 for new participants, designed to cover the programme's international implementation including OECD staff costs and sub-contractors (Bloem 2013). Within this process, participating countries must usually bear the costs related to the national implementation of the ILA: review of items, sampling activities, training of the test instructors, data coding, analysis of national results, material costs such as translation, printing, travel expenses etc. (Wagner 2011; Bloem 2013; Baird et al. 2011). Although in some cases, the involvement of LMICs in

PISA and other ILAs is supported by other international organisations, donor agencies or countries, this is not a given for all participants (Bloem 2013).

National institutional capacity for implementation

The implementation of a national survey according to the standards required by an ILA places substantial demands on the participating national organisations to apply effective methods of oversight, management, organisation and revenue control, as well as the capacity of technical experts and trained personal to design tools, implement surveys and process the data (Bloem 2013; Baird et al. 2011). Many LMICs have yet to develop the wider technical capacities that allow them to undertake activities according to ILA standards without substantial technical and/or resource-based support (Murimba 2005).

Opportunity to gain insight from the results

The national policy benefit of ILA participation is often determined by the extent to which the results can serve national education systems' objectives and interests. For example, in the case of PISA, most OECD countries produce national reports that deal with contextual issues in much greater detail than the international reports. However, among LMICs, without substantial support and guidance, such reports are less frequently produced, and may not have the levels of detailed secondary analysis that will benefit educational policy reform in-country. This can be due to a variety of reasons including low analytical capacity, human and financial resource restrictions, and lack of political interest (e.g., low performance may lead to results being withheld rather than further analysed and disseminated among stakeholders) (Bloem 2013).

Opportunity for full participation and international collaboration

International participation and collaboration, including inter-country exchange of expertise and best-practice, is also presented as a key benefit for participants in ILAs (Lockheed in Wagner et al. 2012). However, the skills required of national co-ordinating bodies and their representatives who can facilitate exchange of best interest to their country context, plus the technical criteria association with their recruitment and the costs of their participation, is often highly demanding (Murimba 2005). Without appropriate levels of support from ILAs themselves, this can result in low levels of international engagement among LMICs, impeding national ownership, motivation and engagement (Bloem 2013).

As partial evidence in support of the above, Best et al.'s (2013) synthesis report notes that, in country cases where it was explicitly stated that LSEAs had delivered no impact on policy process at the national level, the following barriers to the effective use of assessment data were noted:

- problems relating to assessment programme design and analyses
- financial constraints
- weak assessment bodies and fragmented government agencies
- low technical capacity of assessment staff

This infers that if LMICs are to use participation in ILAs to build national capacities to design, implement and make use of national assessments, then the investment in and building of appropriate systemic mechanisms and capacity, in terms of national-level skills, knowledge and financial and technical resources, need to be featured as part of the ILA participation process.

To illustrate the position of international LSEAs in acknowledging and addressing such issues, and to present summaries of evidence of the impact of participation on national capacity development, below are outlines of two approaches to national capacity development, the first from an ILA (PISA/PISA-D) and the second from an RLA (SACMEQ).

ILAs and capacity development: the case of PISA and PISA-D

PISA does not focus on national analytical capacity building, and as such has undertaken no research to assess its impact in this area at the national level. Participating countries may improve their analytical capacity as a result of their participation in PISA, and this is considered by PISA as a positive feature, but it is not regarded as the principal objective (Bloem 2013).

There is evidence among OECD partner countries that policy-makers have utilised the PISA assessment frameworks and instrument as a best-practice model or guide in improving national assessment policies and practices, and have explicitly sought to incorporate PISA-like competencies in revised national standards and curricula. In this context, participation in or exposure to PISA approaches and methodologies might be seen to have contributed to improved capacity of theory and application at a national level (Breakspear 2012).

As part of programme efforts to support LMICs to benefit from this opportunity, PISA-led assistance from OECD countries to support national-level capacity building among LMIC PISA participants has grown in recent years. This includes the development of national reviews under the series *Lessons Learned from PISA* (Bloem 2013). However, in order to help LMICs apply the benefits of PISA participation to developing capacity associated with the design and implementation of NLAs, further support is seen as necessary. Recommendations have included:

- OECD-based technical assistance to support secondary national data analysis;
- Funding to develop assessment materials appropriate to national contexts;
- Establishment of language-group networks for collaborative development of research tools;
- PISA-hosted data analysis workshops for national staff (Bloem 2013)

In response to these and other issues, OECD launched PISA-D in 2013, whose primary mandate is to improve the mechanisms associated with PISA in terms of their contextual relevance and policy focus for LMICs. This will be undertaken in part by developing context-specific data collection tools that take better account of background factors such as socio-economic status and learning environments, and adjusting existing PISA test instruments to a broader range of learner performance levels (OECD 2013ii).

Unlike PISA, PISA-D also explicitly presents country engagement and participation as an opportunity for improving institutional capacity (OECD 2013i; Bloem 2013; OECD 2015i-v). Firstly, OECD states that the initiatives described above are specifically designed to enable more countries to use PISA-based mechanisms in setting national learning targets, monitoring progress, and analysing factors affecting student outcomes – particularly among poor and marginalised populations. Secondly, PISA-D will put in place mechanisms in place to build national capacity by working initially with a group of partner countries over a 36-month period to help to develop enhanced survey instruments and methodologies and undertake field trials and surveys. The results of these will then be used inform new policies.

Finally, in order to support countries to undertake these activities, under PISA-D, OECD supports each partner country in securing financial and technical support from bilateral and multilateral donor agencies, and in forming technical partnerships with networks of experts that will continue to meet on a regular basis (OECD 2013ii).

PISA-D's capacity development procedures are summarised in Annex 1. Examples of the application of PISA-D's approach to national capacity development are provided in the cases of Guatemala and Senegal (Annexes 2 & 3). However, as yet, there is no available evidence of whether these planned interventions, which commenced in Sept 2015, have impacted on capacity for national learning assessment.

RLAs and capacity development: the case of SACMEQ

The cross-national Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) assessment programme, involving 15 regional country participants, does include activities and outputs that focus on national analytical capacity building. The mandate of SACMEQ is to undertake research in order to generate information for policy making, but with a strong focus on training activities to obtain the technical skills required to monitor, evaluate and compare empirical findings for educational planners and researchers within the ministries (Bloem 2013). To this end, SACMEQ's stated mission in fact prioritises technical capacity development above other activities:

> *To undertake integrated research and training activities that will expand opportunities for educational planners and researchers to: (a) receive training in the technical skills required to monitor, evaluate, and compare the general conditions of schooling and the quality of basic education, (b) generate information that can be used by decision-makers to plan the quality of education, and (c) to utilize innovative information dissemination approaches and a range of policy-dialogue activities in order to ensure that SACMEQ research results are widely discussed, debated, and understood by all stakeholders and senior decision-makers and then used as the basis for policy and practice*

In keeping with this remit, SACMEQ have developed a series of training modules in learning assessment design and implementation. These form the basis of individual technical capacity development among educators and civil servants in participating countries. Their delivery is augmented by an ongoing series of action-orientated training workshops that operate to a) facilitate assessment design and implementation activities at both country and regional level, b) address logistical issues, and c) co-ordinate the regional implementation and analysis of each SACMEQ cycle.

In addition, SACMEQ also supports independent higher-level research and publication across the region, thereby assisting in the development of research capacity and secondary data analysis at the national level.

According to Murimba (2005), as a result of such activities, each ministry participating in SACMEQ ministry has reported positively on the benefits of capacity-building. However, this impact has been reported to be especially pronounced in Botswana, Lesotho, Seychelles and Malawi, where in each case, specific capacity needs were identified and addressed through targeted action. Further to this, SACMEQ is seen as being instrumental in enhancing resource capacity, such as through the establishment of ICT (information and communication technology) units for data processing and storage (e.g. in Tanzania), facilities which have subsequently been used for tasks beyond SACMEQ, thus yielding benefits to the entire system (ibid). Finally, from a systemic perspective of capacity development, Murimba also reports evidence of the drive by ministries to institutionalise the SACMEQ process in education systems through the deliberate, planned integration of SACMEQ processes into regular ministry functions, such as routine data collection.

Some further evidence of SACMEQ's impact on capacity development for learning assessment at a national level is available through evaluations of programme input undertaken by key providers (e.g. UNESCO in Ercikan et al. 2008), including reports of high levels of satisfaction among participants in training workshops and other technical support initiatives with the relevance and quality of technical support provided.

6. The impact of participation in ILA on political decisions, policy-making and teaching practices

One of the strengths of ILAs and RLAs is their ability to provide some way of comparing across nations, regions and continents, using the best methodological tools available to generate national summative scores on international tests, and such international comparisons have provided opportunities for policy making and employment of new best practices (Wagner 2011).

Yet, in the context of LMICs, there are frequently challenges to using these opportunities for best advantage, making international comparability of limited importance in securing improvements in learning achievement. Firstly, LMICs participating in inter-country LSEAs frequently make compromises to achieve cross-national consensus, for example by limiting population sampling by excluding marginalised groups and languages (Baird et al. 2011; Wagner 2011; Green & Oates 2007). Secondly, LMICs can find LSEA 'league tables' less useful, where score-based comparisons with high-achieving countries have little value for guiding policy reform (Wagner 2011).

Thirdly, depending on the national context, there can be a range of existing barriers to the effective use of assessment data to inform policy-making in LMICs. Possible factors are related to the nature of the assessment programme itself, the analysis of assessment outcomes, the dissemination of findings from the programme, the nature of the education system and the nature of the political system and wider context. Under these headings, those factors affecting the use of ILA and RLA assessment data at the country level listed by Knight, Lietz, Nugroho and Tobin (2012) include:

- Factors relating to the assessment programme:
    - The appropriateness of the assessment instrument, sampling approach and administration procedures
    - How well the assessment programme is integrated into existing structures, policy and decision-making processes
    - The level of involvement of policy makers in the design and implementation of the assessment programme
- Factors relating to the analysis of assessment outcomes:
    - Whether secondary or in-depth analysis of data is undertaken at country-level
    - Whether analyses focus on diagnosing issues in the education system
    - Whether analyses identifies factors associated with achievement
- Factors relating to the dissemination of findings and analysis:
    - Timeliness of results dissemination
    - Availability of appropriate targeted reports e.g. for senior policy makers, curriculum developers, teachers, the media
    - Whether findings are communicated to meet the needs of specific stakeholders
    - Whether policy makers can interpret findings
    - Whether policy makers place value that is placed on assessment findings, research inputs and evidence in general
- Factors relating to the education system include:
    - The effective functioning of the education system
    - The strength of teachers' unions and their role in policy making
    - Lines of communication and dissemination from decision makers to schools and stakeholders
- Factors relating to wider political system include:
    - The existence of political sensitivities over findings
    - Levels of decentralisation and transparency within the political system

- Levels of public representation in government
- The extent of academic and media freedom, and the strength of civil society
- The strength of public accountability systems
- Existence of conflict or political volatility
- The role of external (e.g. multilateral and bilateral) agencies in the system

The promotion of effective national capacity to conduct learning assessment relies on the extent to which initiatives can identify and address issues associated with the above factors on a contextual basis.

In terms of evidence, Best et al. (2013) state that although almost two-thirds of all developing countries have participated in a national, regional or international assessment programme, little is known about the use of large-scale assessment data in political decisions, policy-making and teaching practice at a national level. At a regional level, there is little available evidence covering Asia, Middle East & North Africa. Research is especially limited for the Pacific – although a recently-published systematic review by Tobin et al. (Sept 2015) examines the link between participation in large-scale assessment programs of students' learning and education policy in 32 countries in the Asia-Pacific region. SACMEQ, covering East and Southern Africa, is seen to have been a relative success in terms of its impact on policy-making, although further investigation is necessary to detail which aspects of the programme exactly contributed to its effectiveness in this regard. However, we will discuss each of these issues in more detail here.

Political implications of ILA participation

According to analysis by Bloem (2013), among LMIC governments, fear of bad performance in ILAs (i.e. appearing at the bottom of the PISA "league tables") is seen as a potential deterrent to participation. The likelihood of low performance may be seen as inviting criticism from education stakeholders and the general public. Furthermore, results may not only reveal low performance, but may also shed light on inequalities within a country which policy makers would not like to make public. Finally, statistics and performance measures increase transparency and demands for accountability, which may not be regarded by governments as being in their interests. In such circumstances, in reaction to disappointing performance in ILAs, certain governments can pay little attention to results and do not make any further analysis of results in the national context.

Murimba (2005) also notes that effective LSEAs are designed to stimulate reform, and thereby have a potentially destabilizing effect on systems. In contexts where high-level officials have initiated, participated in and promoted change, these underlying values support change. However, in systems that have a measure of inertia, change is considered risky. He cites the case of a SACMEQ national regional co-ordinator who received a remote and isolated 'punishment posting' when he reported that, based on SACMEQ results, it was clear that extra tuition among teachers in the education system had reached alarming proportions.

In response to this generalised potential for reluctance to engage with ILAs and reform at a political level, recommendations have included encouraging governments to initially participate in ILAs as "associates," facilitating the gathering of data without requiring international release of national results. This interim arrangement would promote the generation of much-needed data, provide access to expertise, contribute to building local capacity to develop, administer, and analyse tests, yet avoid the political consequences of possible poor performance (Braun et al. 2006).

Evidence of the impact of ILA on political decisions and policy-making

ILAs such as PISA, PIRLS and TIMSS, as well as RLAs such as SACMEQ, LLECE and PASEC, receive considerable media and policy attention, which has led at times to significant international and national educational policy debates and shifts (Wagner 2011). In this

context, at a national level, one of the main aims of conducting ILAs is seen as to provide information on a country's educational outcomes, which in turn, either as part of the official reporting process or resulting from secondary analysis, assists policy-makers and other stakeholders in the education system with making policy and resourcing decisions for improvement (Knight, Lietz, Nugroho & Tobin, 2012).

However, despite this, little is known about how ILAs are used in policy making in LMICs, or about the effects of their basing policy decisions on ILA findings (Kellaghan, Greaney & Murray, 2009; Tobin et al. 2015). As an example of this gap in data, the EPPI-Centre's evidence library houses six systematic reviews on the topic of LSEA assessment, none of which looked at the impact of assessments at the level of the education system, or on the policy making process (Knight, Lietz, Nugroho & Tobin, 2012). Further, Knight et al. (ibid) state that, other than their own activities, no systematic review on the impact of ILAs on policy-making had been completed at the time of writing, whether within the context of LMICs or elsewhere.

Given this, concrete evidence that international LSEAs have had direct policy impact in developing countries is relatively slim, and largely based on non-scientific surveys and interviews (Lockheed in Wagner et al. 2012). But of the available evidence, policy changes associated with the outcomes of ILAs include:

- Resource allocation policies to improve the quality of teachers and teaching materials and increase teacher quality (e.g. through in-service professional development and improved pre-service preparation);
- System-level policies covering curriculum standards and reform, performance standards and assessment;
- Funding policies intended to improve educational outcomes (e.g. interventions and programmes for low-performing schools and those with low-socioeconomic status; performance-based financial incentives for schools and teachers; funding allocation between public and private schooling sectors).
  (Lietz & Tobin 2014; Knight et al. 2012)

Summaries of evidence of the impact of PISA and SACMEQ on political decisions and policy-making are included in Annexes 4 & 5.

Evidence of the impact of ILA on teaching practice

The impact of large-scale assessments on teaching and learning practice policies in LMICs is observed less frequently than impact on policies regarding resource allocations (Best et al. 2013; Fagazollo 2009; Lietz & Tobin 2014; Tobin et al. 2015). Fagazollo (2009) states that the influence of ILAs, and of PISA in particular, is perceived to have impacted on teaching practice through an increasing focus on testing and evaluation – both of pupils and teachers, as well as of schools. However, among LMICs, ILAs most frequently impact on education policies aimed at increasing teacher quality through in-service professional development and improved teacher preparation (Lietz & Tobin 2014; Tobin et al. 2015).

In terms of teaching practice specifically, the primary impact of ILAs is reported to be on approaches and interventions to improving the learning processes by way of student-oriented pedagogy and in-class learning strategies. In terms of frequency, this is followed by an emerging focus on the development and dissemination of targeted teacher and leadership initiatives to assist with improving practice and leadership (Best et al. 2013). However, Murimba (2005) reports that, in the case of SACMEQ countries, evidence shows only Mozambique has used SACMEQ results to directly stimulate innovative pedagogical practices.

7. Consequences of focusing assessment of learning on language, numeracy, maths and science

In general terms, this topic is regarded as both politically and academically contested (see Section 10 Experts' Comments). In addition, discussion of the nature of assessment for learning attainment in specific subjects involves approaches very different from those associated with assessment of learning attainment at a national scale. In addition, some careful distinction is required between learning attainment in language and that in reading, and also between numeracy and maths.

Primarily, the focus on these core subjects is regarded as a recognition of the important foundation they provide for all school learning. Children who do not learn to read in the first few grades are likely to repeat grades and to drop out at an early stage (Kellaghan, Bethell, & Ross, 2011). A focus on basic skills is also regarded as contributing to core skills and knowledge associated with individual economic autonomy (*The Economist*, 2013).

However, the association that ILAs such as PISA make between national attainment in these subjects and domestic economic growth is seen as 'distorting educational policies' and encouraging governments to adopt an overly 'economic' approach to education (Froese-Germain 2010; Morrison 2009). Combined with the use of league tables as the primary tool of international comparison, the effect of ILAs has been likened to 'Education Olympics' (Froese-Germain 2013) in which economic competition against other countries becomes the goal of education policy, and of policy reform (*The Economist*, 2013).

Criticisms have also been raised about the disproportionate focus – both in the form of policy and resources – on such a narrow range of subjects, at the expense of Humanities and the Arts, where curriculum time can be reduced in order to dedicate more hours to numeracy and literacy (People for Education, 2013). For some commentators, the focus on maths, science and reading among national education systems shifts the attention of governments and policy-makers almost exclusively to these core subjects, leaving subjects such as history, geography, civics, languages and others marginalised, and dividing the curriculum into a core and a less important periphery of all other subjects (Morrison 2009).

From a pedagogic perspective, it has also been claimed that focusing on literacy/numeracy/science is an over-simplistic measure of educational quality which neglects equally important, although arguably more difficult to measure, areas such as creativity, citizenship and mental and physical wellbeing (Froese-Germain 2013). Although an increased focus on targeted subjects can lead to improved results, it is not clear whether that is due to the increased subject time, rote learning to tests and whether achievement in other subjects suffer as a result (Mons 2009).

8. Circumstances and actions required to ensure learning assessments promote and secure improvements in learning achievement

Under the appropriate conditions, both national and international assessments can contribute to quality improvement in learning achievement in LMICs. Findings can provide transparent information about system performance that goes beyond input measures and reveals trends over time. Background information may allow results to be disaggregated for sufficiently large sub-populations, revealing relationships between achievement and student or school characteristics. Such information can guide investment of resources in low-performing geographical, administrative or curriculum areas, or for specific socioeconomic groups (Kellaghan, Bethell, & Ross, 2011).

However, applying this information to improve the quality of student learning and secure improvements in learning achievement involves a commitment to successive assessment exercises over time, which in its turn requires institutionalisation of the assessment process, integration of assessment information into Education Management Information Systems, and an effective series of evidence-based policy reforms directed at school resourcing, classroom practice and teacher development (Kellaghan, Bethell, & Ross, 2011; EQ Review 2010).

While there are questions overhanging the value of using ILAs to compare experiences between developed and developing countries, Bernbaum and Schuh Moore (2012), in their comparison on the use of learning assessments across five countries or states (Brazil, Germany, Namibia, Singapore, Massachusetts USA) conclude that their findings suggest that the actual tools that effect change in improving learning attainment are similar across countries, regardless of the level of development.

In support of this, they cite examples from a range of educational innovations introduced by the governments of Namibia and Brazil benchmarked against successful experiences in developed countries. These include: adopting standards to serve as the basis for developing curricula, textbooks and teaching methodologies; designing assessments to assess whether these standards are being met; and taking steps to increase the quality and professionalism of the teaching force.

However, Bernbaum and Schuh Moore also identify that, while the assessment tools themselves are universally applicable, their ability to positively influence improvement in learning attainment are defined by differences in context between countries. In light of this, the authors identify common contextual factors in learning assessment design, implementation and analysis across the five cases to which they attribute national increases in learning achievement. These are as follows.

Firstly, in general terms, they record a broad appreciation at government level that improving student learning requires a holistic systems approach that addresses a wide range of factors that need to be present for enhancing student learning (including enhanced policies, changes in administrative systems, establishing standards and preparing curricula, teaching according to those standards, training teachers and working with parents). They further noted that, while it is possible for learning assessment reform to improve learning achievement, their cases demonstrated that any reform initiatives need to be long-term and appropriately contextualised.

Secondly, further to both these points, each of their chosen cases demonstrated the following systemic commonalities with regard to learning assessment:

- Continuity in political will, including a shared vision for what the education system should be achieving
- Prioritising equitable access to quality learning as a key element of reforms
- Establishing and implementing clear standards and curricula at each level of schooling
- Ensuring appropriate focus on enhancing student performance in math and science
- Recruiting highly qualified individuals to the teaching profession
- Maintaining high-quality teacher professional development programmes at pre-service and in-service
- Establishing systems of accountability for schools and standards

Case studies from Brazil and Uganda, two countries active in addressing such capacity level concerns, are included in Annexes 6 & 7.

Further commentators highlight the importance of specific school-level assessment initiatives in facilitating improvements in learning achievement.

Firstly, Kellaghan, Bethell, and Ross (2011) and Green and Oates (2007) claim that for assessment to support improvements in learning achievement, there has to be: high quality information on trends in attainment; support for school improvement processes; and clear application of valid assessment models directed at enhancing learning attainment.

Secondly, within this, the design of learning assessments to support the enhancement of learning achievement should include specific mechanisms to deliver relevant information on learning attainment to pupils, parents and teachers to enhance learning, operate systems of school-level accountability for schools, and deliver highly robust school-level information for policy purposes (Green & Oates 2007). At its most basic level, this involves placing assessment design at the classroom level, e.g. through formative, summative and diagnostic tests which can help teachers and the same time as providing information for policy and planning at central and decentralised levels (EQ Review 2010).

Green and Oates (2007) also outline a range of possible models for national-level assessment to improve learning achievement, illustrating the fact that there are a number of radically different possibilities that are worthy of consideration. They further highlight the importance of context-specific research and development in the design and application of assessments for learning attainment, including rigorous piloting and evaluation, to ensure the suitability of such tools at the classroom level. Any effective national assessment system for improvement of learning attainment should be designed specifically to be fit for purpose and supportive of teaching and learning.

In such environments, there are general concerns about the role that ILAs and RLAs can play. In addition to logistical concerns surrounding the frequency of LSEAs and the time involved in undertaking secondary and (at classroom level) tertiary analysis, Green and Oates (2007) highlight the disparity between the forms of data required by government at regional and international levels and the forms of teacher-led assessment to carry out for teaching and learning purposes. They consider the use of national test data for multiple purposes as having a negative impact, and such issues of disparity of data have to be considered seriously as part of any investigation into learning assessment for learning achievement.

9. References

Baird, J. et al., (2011), Policy Effects of PISA, Oxford Centre for Education Assessment, http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/10/Policy-Effects-of-PISA-OUCEA.pdf

Bernbaum, M. & Schuh Moore, A. (2012) Examining the Role of International Achievement Tests in Education Policy Reform: National Education Reform and Student Learning in Five Countries, EQUIP2, USAID

Best, M. et al., (2013) Systematic Review: The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries, AusAID, https://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=KKB8wH9tvmA%3D&tabid=3418

Bloem, S. (2013), "PISA in Low and Middle Income Countries", OECD Education Working Papers, No. 93, OECD Publishing. http://dx.doi.org/10.1787/5k41tm2gx2vd-en

Braun et al., (eds.) (2006), Improving education through assessment, innovation and evaluation, Cambridge, MA: American Academy of Arts and Sciences.

Braun, H., and Kanjee, A. (2007) Using assessment to improve education in developing nations. In Educating all children: A global agenda, ed. J.E. Cohen, D.E. Bloom, and M.B. Malin, 303–53. Cambridge, MA: MIT Press.

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, OECD Publishing. http://dx.doi.org/10.1787/5k9fdfqffr28-en

Chulu, B.W. (2009), Institutionalisation of assessment capacity in developing nations: The case of Malawi. Paper presented at the 3rd global meeting of the Russian Aid for Education Development programme, 30 September–2 October, in Kenya.

Clarke, M. (2011), Framework for Building an Effective Student Assessment System, READ/SABER Working Paper, accessed at: http://files.eric.ed.gov/fulltext/ED553178.pdf

Crooks, T. (2002) Educational assessment in New Zealand schools. Assessment in Education 9, no. 2: 217–37.

Darling-Hammond, L., and Wentworth. L. (2010) Benchmarking Learning Systems: Student Performance Assessment in International Context. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

The Economist (2013) 'PISA Envy' http://www.economist.com/news/international/21569689-research-comparing-educational-achievement-between-countries-growing-drawing

Ercikan, K. et al. (2008), Evaluation of Dimensions of the Work of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and of its Programme of Cooperation with the International Institute for Educational Planning (IIEP) , UNESCO http://unesdoc.unesco.org/images/0016/001626/162675E.pdf

EQ Review (2010) Using Learning Standards to Promote Student Achievement, EQ Review, Vol 8, No. 1 http://www.equip123.net/EQ_Review/8_1.pdf

Ferrer, Guillermo (2006). Educational Assessment Systems in Latin America: Current Practice and Future Challenges, Washington, DC: PREAL.

Figazollo, L. (2009). 'Impact of PISA 2006 on the education policy debate'. Brussels: Education International.

Froese-Germain, B. (2010) The OECD, PISA and the Impacts on Educational Policy by CTF Researcher, http://files.eric.ed.gov/fulltext/ED532562.pdf

Froese-Germain, B. (2013) Commentary on the OECD's Programme for International Student Assessment 2012, CTF Research and Information http://www.ctf-fce.ca/Research-Library/PISA2012commentary.pdf

Green, S. and Oates, T. (2007) How to promote educational quality through national assessment systems, Cambridge Assessment. A paper presented at the International Association for Educational Assessment Annual Conference, Baku, Azerbaijan, http://www.cambridgeassessment.org.uk/images/109749-how-to-promote-educational-quality-through-national-assessment-systems.pdf

Grisay, A., de Jong, John H.A.L., Gebhardt, E., Berezner, A., Halleux-Monseur, B. (2007) Translation Equivalence across PISA Countries. Journal of Applied Measurement, 8 (3), 249 – 266 https://www.acer.edu.au/files/grisay.pdf

Kanjee, A. (2008), A framework for developing national education assessment systems: Challenges and prospects. Paper presented at the National Policy dialogue forum on developing effective assessment systems, November 19–21, in Pretoria.

Kanjee, A., and Acana, S. (2010). Creating an enabling environment for an effective assessment system in Uganda. Paper presented at the 2nd global meeting of the Russian Aid for Education Development programme, October 4–6, in Kenya.

Kellaghan, T., Bethell, G. & Ross, J. (2011) National and international assessments of student achievement, A DfID Practice Paper, DFID https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67619/nat-int-assess-stdnt-ach.pdf

Kellaghan, T., Greaney, V., Murray, T. S. (2009) Using the Results of a National Assessment of Educational Achievement. World Bank. https://openknowledge.worldbank.org/handle/10986/2667

Knight P, Lietz P, Nugroho D, Tobin M (2012) The Impact of National and International Assessment Programmes on Educational Policy, Particularly Policies Regarding Resource Allocation and Teaching and Learning Practices in Developing Countries. Protocol. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Larach, L., and Lockheed, M. (1992) "World Bank Lending for Educational Testing." PHREE Background Paper, 92/62R. Population and Human Resources Department. Washington, DC: World Bank.

Lietz, P. & Tobin, M. (2014) "Informing policy in developing countries", International Developments: Vol. 4, Article 6. http://research.acer.edu.au/intdev/vol4/iss4/6

Mons, N. (2009), Theoretical and Real Effects of Standardised Assessment, Eurydice Network http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111en.pdf

Mortimore, P. (2009) Alternative Models for Analysing and Representing Countries' Performance in PISA. Brussels www.ei-ie.org/research/en/documentation.php

Murimba, S. (2005). The impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). Prospects , XXXV (1), pp. 91-108. http://www.springerlink.com/content/7126vx7082543x83/

Nzomo, J. and Makuwa, D. (2006). How can countries move from cross-national research results to dissemination, and then to policy reform? In K. N. Ross and I. Jurgens-Genevois (Eds.) Cross-national studies of the quality of education: Planning their design and managing their impact (pp. 213-228). Paris: IIEP (UNESCO) http://www.unesco.org/iiep/PDF/pubs/G117.pdf?class=IIEP_PDF_enhp&page=G117&estat_url=http://www.unesco.org/iiep/PDF/pubs/G117.pdf

Oduol, T. (2006). Towards the making of education policy in Kenya. International Education Journal, 7(4), 466-499. http://ehlt.flinders.edu.au/education/iej/articles/v7n4/Oduol/paper.pdf

OECD (2013i), The OECD's contribution on education to the post-2015 framework: PISA for Development introduction http://www.oecd.org/pisa/aboutpisa/POST-2015%20%20paper%20education_20%20Feb.pdf

OECD (2013ii), Improving Learning Outcomes Worldwide: How PISA can help http://www.oecd.org/pisa/aboutpisa/pisa-for-development-brochure.pdf

OECD (2015i), PISA for Development, Guatemala: Capacity Needs Analysis http://www.oecd.org/pisa/aboutpisa/NEW_Guatemala%20CNA%20reportFINAL_SL.pdf

OECD (2015ii), PISA for Development, Guatemala: Capacity Building Plan http://www.oecd.org/pisa/aboutpisa/Guatemala%20CBP%20report_FINAL2.pdf

OECD (2015iii), PISA for Development, Senegal: Capacity Needs Analysis http://www.oecd.org/pisa/aboutpisa/NEW_Pisa-D_CNA_Senegal_FINAL_revised.pdf

OECD (2015iv), PISA for Development, Senegal: Capacity Building Plan http://www.oecd.org/pisa/aboutpisa/Senegal%20CBP%20Report_FINAL.pdf

OECD (2015v), PISA for Development, Zambia: Capacity Needs Analysis http://www.oecd.org/pisa/aboutpisa/NEW_Pisa%20for%20Development_Zambia_FINAL_revised.pdf

People for Education (2013), Broader measures of success: Measuring what matters in education http://www.peopleforeducation.ca/measuring-what-matters/wp-content/uploads/2013/10/P4E-MWM-full-report-2013.pdf

Ravela, P. (2005) "A Formative Approach to National Assessments: The Case of Uruguay." Prospects 35(1): 21-43.

Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R., & Nyamkhuu, T. (2015). Using large-scale assessments of students' learning to inform education policy: Insights from the Asia-Pacific region. Melbourne: ACER and Bangkok: UNESCO.

Wagner, D. (2011) Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries, UNESCO, Paris http://unesdoc.unesco.org/images/0021/002136/213663e.pdf

Wagner, D., Lockheed, M., Mullis, I., Martin, M.O., Kanjee, A., Gove, A. & Dowd A.J (2012): The debate on learning assessments in developing countries, Compare: A Journal of

Comparative and International Education, 42:3, 509-545
http://dx.doi.org/10.1080/03057925.2012.670480

Wasanga, P. (2010) The role of examinations results as a feedback tool: The Kenyan experience. Paper presented at the 2nd Global meeting of the Russian Aid for Education Development programme, October 4–6, in Kenya.

10. Additional information

Authors
This query response was prepared by **Chris Joynes (CfBT)** and **Helen Mobey (CfBT).**

About Helpdesk reports: The HEART Helpdesk is funded by the DFID Human Development Group. Helpdesk reports are based on 3 days of desk-based research per query and are designed to provide a brief overview of the key issues, and a summary of some of the best literature available. Experts may be contacted during the course of the research, and those able to provide input within the short time-frame are acknowledged.

For any further request or enquiry, contact info@heart-resources.org

HEART Helpdesk reports are published online at www.heart-resources.org

ANNEX 1: Summary of PISA-D's capacity development procedures

In support of achieving the above, PISA-D has a series of capacity development exercises that it will undertake with all country participants. These are based on the execution of a Capacity Needs Analysis and the subsequent development of a Capacity Building Plan. Within this, country capacity needs are distinguished according to a) those required to undertake PISA-D, and b) those required to implement national learning assessments (NLAs) (OECD 2015i-v). As such, this represents PISA-D's specific commitment to enabling the development of country capacity to undertake NLAs.

PISA-D's capacity development framework is structured to undertake needs analysis and building plans according to three dimensions: 1) capacity within the enabling context, 2) organisational capacity and 3) individual capacity. Within this, the framework is primarily designed to assess country capacity in relation to the five PISA-D programme outputs, covering development of:

- 'enhanced contextual questionnaires and data-collection instruments;
- enhanced descriptive power of cognitive assessments in reading, mathematics and science, at appropriate skill levels within the PISA cognitive framework;
- an approach, including a methodology and analytical framework, for including out-of-school 15-year-olds in the assessments;
- increased country capacity in assessment, analysis and use of results for monitoring and improvement;
- engagement with OECD, development partners and other developing countries in order to identify peer-to-peer learning opportunities regarding participation in PISA and its potential contribution to the UN-led discussions on the post-2015 framework.' (OECD 2015i-v)

Examples of the application of PISA-D's approach to national capacity development are provided in the cases of Guatemala and Senegal (Annexes 2 & 3). However, while the Capacity Needs Analyses and Capacity Building Plans lay out plans for the development of national capacity in each case, the interventions focus on prioritising the implementation of PISA-D. In each case, the plans make only incidental reference to the strengthening national learning assessment mechanisms, which will occur from association with the PISA-D interventions.

ANNEX 2: PISA-D capacity development in practice: Guatemala

Contextual analysis

Undertaken between 2014 and 2015, PISA-D's capacity needs analysis concluded that Guatemala is well-positioned to begin preparation for implementation of PISA for Development. The country was assessed to already have a strong, well-established and well-managed national assessment programme, run since 2006 and drawing in part on its previous participation in other ILAs (i.e. SERCE, TERCE, ICCS).

At government level, Guatemala has a dedicated assessment unit (DIGEDUCA), whose members were assessed to be well-trained, 'hugely committed', and operating on a firm basis. In addition, the unit is institutionally enabled to allow for continuing training. More widely, further national capacity is represented by the commitment to research shown within higher education, specifically Del Valle University's Faculty of Education and its Masters programme in Education Research, Assessment and Measurement programme.

Finally, Guatemala publishes a yearly series of reports regarding annual tests, demonstrating the country's capacity to gather, analyse, and disseminate data on a regular basis (OECD 2015i).

Recommendations

The above context notwithstanding, PISA-D's analyses indicated a number of areas where national capacity could be enhanced. In summary, the priorities for each of the three PISA-led dimensions (enabling environment, organisation and individual) for Guatemala are set out below:

- Enabling environment:
  - Ensuring funding for specific PISA activities, e.g. international travel for meetings
  - Using reports to better establish the relationship between assessment results and policy measures
  - Ensuring stakeholder access, understanding and application of result reports
- Organisational capacity:
  - Enhancing national capacity in assessment, analysis and use of results for monitoring and improvement of education quality by including open questions in tests
  - Enhancing cognitive assessments for below-baseline proficiency levels in PISA
  - Enhancing contextual questionnaires and data-collection instruments
- Individual capacity:
  - Enhancing abilities in developing open questions in tests
  - Enhancing abilities in cognitive assessments for low-level proficiency levels in PISA (OECD 2015i)

Furthermore, PISA-D's analyses indicated a number of areas where national capacity could be enhanced in relation to a) implementation of PISA-D and b) general enhancement of assessment capabilities, including for development of national learning assessment mechanisms.

For implementation of PISA-D, the areas for capacity building included:

- developing the commitment of stakeholders to PISA-D
- developing new quality procedures and enhancing those established for the implementation of PISA-D

- enhancing procedures for sampling schools, students and out-of-school young people
- enhancing research products, communication and dissemination.

For general enhancement of assessment capabilities, the areas for capacity building included:

- improving assessment infrastructure
- enhancing psychometric test methods
- improving preparation and storage of test items
- enhancing research methodologies.

(OECD 2015ii)

In addressing these identified capacity development needs, PISA-D has set out in the Capacity Building Plan a series of costed initiatives, primarily activity-focussed meetings and training workshops, scheduled to take place between Sept 2015 – Dec 2018. The priority for these activities focuses on the implementation of PISA-D, and the plan makes only incidental reference to interventions for strengthening national learning assessment mechanisms. The total budget for these activities is $204,500. While the government will cover this cost internally, the majority of the budget associated with these initiatives is for travel to take part in international workshops. As the State of Guatemala does not fund international travel for officials, PISA-D is facilitating discussions between Guatemala and GIZ to provide financial support in this area (OECD 2015ii).

As yet, there is no evidence of the impact of these measures on enhancing capacity for national learning assessment within Guatemala.

ANNEX 3: PISA-D capacity development in practice: Senegal

Contextual analysis

PISA-D's capacity needs analysis concluded that, overall, Senegal has a solid foundational capacity for implementing PISA-D. Senegal conducts both national and international assessments on a regular basis: it has conducted since six waves of national assessment (Système National d'Évaluation de Rendement Scolaire) since 1994, and has participated twice (2007 and 2013) in PASEC. These assessments were carried out by the Division of Evaluation of INEADE (Institut National d'Étude et d'Action pour le Développement de l'Éducation), which will also implement PISA.

INEADE has experience with the design and implementation of traditional assessments, constructing cognitive instruments and contextual questionnaires, translating them from French to Wolof, and conducted assessments on randomly selected children. INEADE has also led the training and co-ordination of data collectors, and the establishment and maintenance of protocols for data protection and confidentiality.

INEADE is aware that engagement with PISA-D will require sufficient staff with multidisciplinary skills, and has set up a PISA-D team of 10 people mostly drawn from the Division of Evaluation and the Division of Studies and Curricula. This team have worked with context questionnaire design but need to familiarise themselves with PISA methodologies.

In general, PISA-D felt that Senegal meets the established category requirements for participation in PISA. It was felt that some procedural improvements can be made by INEADE using models of international best practice without specific training. However, many of the outstanding capacity needs were found in association with supporting the more complex requirements for PISA and the technical and logistical capacities that they require. Specifically, these include the following:

- The Senegal PISA team require training to develop a deep understanding of the PISA framework and its use in the development of instruments and the interpretation of results. They will also require training on PISA tools and methodologies. Finally, the team will then need to explain these to ministry, administrators, teachers, researchers and other stakeholders
- Translators, test administrators and coders will require training on PISA frameworks in French, Wolof and Arabic, and all PISA-related training manuals and instruments will require translation.
- Development of protocols on conflict of interest, confidentiality, data access, tests of reliability and pre-testing, to be provided in French, Wolof and Arabic: this will be the first time that data will be collected in Arabic by INEADE.
- Sampling frames for in-school and out-of-school children will need to be developed in 3 languages, together with assessment manuals and procedures for out-of-school children.
- Thus far, the link between assessments and policy reform has been weak. Channels for feeding evidence into policy-change need to be systematic and institutionalised, including actively linking to the policy cycle of government and school administration to ensure results are used, systematic provision of ministerial briefs, and general timeliness in dissemination.
- INEADE require a strong communication plan to garner media attention and explain the complexities of PISA and the country's ranking and standing in the domains to a broad variety of audiences. Thus far, there has been little national communication regarding assessment results.

(OECD 2015iii)

Recommendations

PISA-D see Senegal's participation requiring the development of financial, human resource and technical infrastructure capacity in order to draw the most out of both PISA and its own national assessments. In the first instance, Senegal may benefit from contact with PISA countries using French and Arabic.

Outside of this, the priorities for each of the three PISA-led dimensions (enabling environment, organisation and individual) for Senegal are set out below:

- Enabling environment:
    - Development of functional plans to facilitate PISA implementation, including
        - plans for developing preparatory materials about PISA for schools, teachers, inspectors, principals and teacher training;
        - schedule of PISA-related research products produced in relation to the policy and planning cycle of government and for key target audiences.
    - Development of a multi-source PISA budget framework, detailing direction of donor funds to specific purposes
    - Development of a PISA stakeholder management plan to ensure support, synergy and complementarity of activities and findings
    - Planning of human resources for PISA implementation period, including required competencies
- Organisation:
    - Improvement of quality standards related to PISA implementation, covering protocols for security, data storage and access to data
    - Upgrading of computers, servers and access to internet to gather and process data
    - Training and preparing of ancillary staff such as data collectors, monitors and coders
    - Development of standard training manuals, coding guides and training modules in the languages of implementation.
    - Implement system for internal and external secondary research using PISA data
- Individual:
    - Training for technical competency on the use of PISA frameworks for instrument and materials design and development
    - Training for analysis and interpretation of results for Senegal

In addressing these identified capacity development needs, PISA-D has set out in the Senegal Capacity Building Plan a series of costed initiatives, structured into an input schedule covering Sept 2015-Dec 2018. These cover a broad range of activities, including: a Strategic Plan for integration of PISA into existing national assessment strategies; Research Development and Communications plans; cycles of recruitment and specialist training across the programme hierarchy; infrastructure investment covering facilities and ICT; and the establishment of steering committees for each area. The priority for these activities focuses on the implementation of PISA-D, and the plan makes only incidental reference to interventions for strengthening national learning assessment mechanisms.

Within this plan, cost items include technical experts, material resources, training implementation costs, production and distribution costs, and travel and subsistence. The total budget for these activities is $460,550. PISA-D is facilitating discussions between INEADE, the Government of Senegal and The World Bank to provide financial support in this area (OECD 2015iv).

As yet, there is no evidence of the impact of these measures on enhancing capacity for national learning assessment within Senegal.

ANNEX 4: ILAs and policy-making: the case of PISA

PISA is differentiated from other ILAs such as TIMMS and PIRLS though its policy orientation. PISA is regarded not only as research, because it also contains key policy recommendations, thereby making it a powerful tool for political influence (Froese-Germain 2010). Critics see PISA as helping governments advocate for 'neo-liberal reform' of national education system, particularly through an efficiency- and competitiveness-driven approach to evaluation and ranking of educational achievement rather than a focus identifying capacity needs and provisions for resources (Fagazollo 2009).

Among OECD member countries, there is strong evidence that PISA is an influential element of education policymaking processes at the national level (Breakspear 2012). There is also preliminary evidence that PISA-based approaches are being integrated within their national policy and practices associated with assessment and evaluation, curriculum standards and performance targets. In particular, PISA-led approaches to assessment of reading, mathematics and science have led to numerous changes to policy at national level, although it is also noted that national findings associated with 'student, interest, engagement, motivation and attitudes' have not generally led to changes in educational policy and practice (Baird et al. 2011; Breakspear 2012). However, although among those member countries with well-developed assessment systems, PISA results are used as one of multiple indicators to inform policy decisions and development (Breakspear 2012).

Also among OECD member countries, there is preliminary evidence that participation in PISA has increased exposure to models of practice among other high-performing national systems, which has in turn been influential on internal policy-making processes. For example, Finland's educational policies, particularly in teacher recruitment and training, school financing and school-based autonomy over curriculum development, are regarded as a key source for cross-national policy learning and borrowing (ibid).

In relation to LMICs specifically, under the PISA-D initiative and its country-associated capacity needs analyses and building plans, there is some evidence that PISA is working to ensure that the appropriate mechanisms are in place for governments to use the assessment and its results to inform policy in an effective and evidence-based manner. However, there is as yet no available evidence on the extent to which policy-level impact has been affected.

ANNEX 5: RLAs and policy-making: the case of SACMEQ

Over the course of the four SACMEQ regional assessment cycles, commenced in 1995, the programme has generated an increasingly sophisticated series of sector-level briefings at both national and regional level, each designed to inform policy decision-making. As the various cycles of SACMEQ have focused in increasing detail on analysing specific areas of data, the briefings themselves have evolved from the generalist 'Some policy suggestions based on a survey of schools' (e.g. SACMEQ I, 2001) to a series of detailed analyses covering the quality of schooling outputs, progress towards gender equity, and school-level awareness of HIV/AIDS as well as subject-specific achievement levels in Maths, Science and Literacy, each making specific policy recommendations (e.g. SACMEQ III, 2011). In addition to national-level reports, these publications include series of working papers comparing national data on specific sub-sectors across the region, and their overall aim is to inform policy-making nationally and regionally.

Regarding this series of publications, Murimba (2005) notes that perhaps the greatest impact SACMEQ has had on ministries of education is on policy-related system development processes. Among countries participating in SACMEQ, by 2006 evidence of those areas impacted by SACMEQ-led policy reform at country level included:

- standards or norms for resource inputs to schools
- policy reform to address grade repetition and extra tuition
- procedures for development planning, policy review and educational reform processes
- policy to ensure equitable access
- quality (by ensuring that pupils mastered basic competencies in English, mathematics, science and skills-related subjects)
- strengthening teacher education and support;
- provision of physical facilities at school level
- focus on HIV/AIDS initiatives
- initiatives for improved achievement in EFA and UPE
  (Murimba 2005; Nzono & Makuwa 2006; Oduol 2006)

Some further evidence of SACMEQ's impact on policy reform at a national level is available through internal evaluations of programme input undertaken by key support providers (e.g. UNESCO in Ercikan et al. 2008). Their findings include documented evidence of national policy change as a result of SACMEQ-led inputs (e.g. Malawi 2004 policy reform p35). However, there is as yet no available evidence systematically assessing the extent to which the range of Policy Briefings and other SACMEQ-led activities have attained policy-level impact across the region.

ANNEX 6: Actions undertaken to ensure learning assessments promote and secure improvements in learning achievement – BRAZIL

Over the past 20 years, Brazil has developed a robust assessment system covering institutionalised exams, large-scale and school-based assessments. Based on its experience, the following lessons can be learnt:

- Developing a national assessment system is a long-term project that must be gradually implemented. There is a long journey from the first assessment efforts (usually once-off assessment exercises) to the establishment of a stable assessment system. For Brazil, this journey took at least 15 years

- Ensuring political leadership. The political engagement of presidents, governors, ministers, and local authorities is critical for the success of an assessment system. This engagement must transcend political coalitions and ideologies, and must be stable in time. Direct connections with the Minister of Education and other stakeholders are strategic to ensure support for assessment programs. These conditions allow for deploying long-term assessment policies. In Brazil, the direct support of presidents Cardoso and Lula allowed for the adoption of a new policy framework, the revamping of an old assessment institute, and the implementation of federal and state assessment programs.

- Establishing a clear policy framework that ensures the stability of assessment activities. In Brazil, the legal framework for assessment provided the architecture for the system. It defined the assessment programs to be implemented and the type of institution that should be in charge of them. It also provided the required funding.

- Designating a clear organization in charge of the assessment. The creation of an independent agency in charge of national assessments, educational statistics, and research is paramount. In Brazil, the designation and revamping of an old assessment institute (INEP) was strategic for the development of the assessment system.

- Avoiding the risk of overloading the leading assessment institution with operational tasks related to assessment implementation. The implementation tasks (test production, distribution, administration, data processing, and reporting) can easily consume all of the time and energy of an institution. However, this should not be the role of a strategic government agency. In Brazil, the federal assessment institute (INEP) has not yet found the right balance between policy/strategy versus implementation. This was very clear during the introduction of the national university entrance examinations. Nowadays, one of the biggest difficulties that INEP is facing is the implementation of the university entrance examination. Since 2009, INEP has been in successive crises because of the politics and the technical challenges of this university entrance examination. As a result, INEP is losing its identity as a leading assessment organization. Instead, it is more and more perceived as the national agency in charge of selecting students into tertiary education.

- Ensuring participation of key stakeholders in running the assessment. INEP outsources key implementation tasks to private organizations. These tasks include training teachers in item writing (enabling them to submit questions to be included in the national assessments) and training test supervisors and administrators in the states and municipalities. This approach allows INEP to focus more on strategy and less on operational tasks (although more needs to be done here). Most importantly, it allows for building capacity and support for assessment all over the country.

- Defining national standards and a curriculum that provides focus to the assessment system. Ideally, this should be done before designing the assessment so that everything is aligned. In Brazil, the introduction of the national university entrance examination followed the introduction of a curriculum reform.

- Building staff capacity to run assessment programs. Options include on-the-job training, and training to enable temporary staff to participate in assessment activities (such as item writing and test administration). Other options include funding for personnel to study abroad in advanced assessment programs (at the masters or doctoral level), or internships in assessment institutions. Brazil has explored all of these options.

- Creating new institutions specialised in student assessment. In 1998, with the support of the World Bank, INEP funded the creation of five assessment centers linked to public universities. Some of these centers have had excellent performance and trained many assessment professionals. In 2001, the creation of the National Association of Educational Evaluation (ABAVE) brought new incentives and support for masters and doctoral studies in assessment. Today, Brazil has many centers of research specialised in student assessment and, most importantly, it has a critical mass of people with expertise in this field.

- The use of assessment results for school accountability purposes may greatly contribute to consolidating an assessment culture. In Brazil, the public dissemination of a school quality indicator (IDEB) fostered a demand for information, and attracted the attention of civil society and the media. This indicator was widely adopted by municipalities and states to monitor and support school performance and to distribute monetary incentives.

- The use of assessment results to inform policy also contributed to consolidating the assessment system, especially among researchers and policy makers. In Brazil, economists and statisticians greatly contributed to identifying factors related to student performance. This information has been regularly used to inform and design education policy.

- A key challenge for consolidating the assessment system seems to be the use of assessment results to inform teachers and pedagogy. There is currently a disconnect between the assessment programs and teacher training institutions in Brazil. This severely impairs the capacity of assessment information to improve teaching practices and learning.

- Countries have to ensure that assessment information is used to improve pedagogy and learning. One mechanism for doing this is by creating assessment programs that have the primary objective of supporting student learning. Brazil did this with its school-based assessment program, Provinha Brasil. Other mechanisms (not yet explored by Brazil) include funding pedagogical research using assessment data, and publishing pedagogical material based on the assessment findings.

- Introducing national university selection examinations is an extremely complex and long-term task. Factors that may facilitate the process are a clear policy document backing the examinations, the voluntary adoption of the system by universities, and the involvement of stakeholders in defining what should be measured. Public trust needs to be built, and the best way to do so is by avoiding mistakes in examination implementation. In Brazil, the full implementation and adoption of the national examinations took more than a decade. It is extremely hard to convince universities to risk their reputation in a national competition for the best students. It is also very

hard to convince them to delegate to a federal agency their power in administering their own admission process.

Source:
Guimaraes de Castro, Maria Helena. 2012. Developing the enabling context for student assessment in Brazil. Systems Approach for Better Education Results (SABER) student assessment working paper no. 7; Washington, DC: World Bank. http://documents.worldbank.org/curated/en/2012/12/17191757/developing-enabling-context-student-assessment-brazil

ANNEX 7: Actions undertaken to ensure learning assessments promote and secure improvements in learning achievement – UGANDA

Uganda's vision of a quality education system with a focus on student learning created the push to strengthen the country's system of assessment activities. As part of Uganda's journey in developing and reinforcing the enabling context for their assessment system, some of the key lessons to be learned include the following:

- Political stability. This allows governments to sustain their commitment to key educational policies and provide consistent leadership at necessary institutions. Political stability also allows for long-term planning and development of various assessments.

- Commitment to global education policies and goals. The Education for All and Millennium Development Goals initiatives, and their like, can act as catalysts for assessment reforms in many countries. The need for planning and monitoring, and for ensuring that all students are learning, has created urgency for strengthening assessment systems.

- Strong policy framework. Policy documents calling for improvements in education and for assessment reforms provide an institutional base for developing stronger assessment systems.

- One central assessment institution. Countries aiming to develop their assessment system may consider having one leading assessment institution. This has several benefits. It reduces wasteful duplication of fiscal, physical, and human resources; streamlines communication between coordinating bodies; and facilitates cross-fertilization of skills and knowledge among staff involved with various assessments. A single institution also can provide a clearer organizational structure and direct accountability for various assessments. Additionally, a single organization facilitates the alignment, compatibility, and synergy between the various assessments, as well as between assessments and curriculum.

- Capacity building. The success of any assessment activity is directly related to the capacity of the personnel developing and implementing it. In developing countries, assessment institutions need to attract personnel who are eager to learn. Opportunities for learning would come from on-the-job experience, from participating in international assessment programs, and from formal training provided by international donors. Competitive salaries are required to keep trained personnel in the assessment institution.

- Institutionalised use of assessment data. Due to limited access to secondary education in Uganda, the public has come to accept the use of end-of-level examinations as the most democratic way to select students for further schooling. In addition, data from the national large-scale assessment, NAPE, have been usefully employed to identify trends in student performance so as to develop interventions to raise the standards of student learning. NAPE data also have been used to address, and verify, public perceptions of reduced quality after implementation of Universal Primary Education. The demonstrated value of these data to inform the government and general population has both validated and institutionalised these assessment activities and highlighted their utility in educational planning and monitoring.

- Curriculum reform. Curriculum reform is an important catalyst for assessment reform. The development of the 2007 Thematic Curriculum provided a strategic opportunity for implementing Uganda's continuous assessment program. Collaboration between

the National Curriculum Development Centre and UNEB in developing aligned and integrated curriculum and assessment activities provided a more structured way to assist teachers in enhancing learner performance. Because curriculum and continuous assessment go hand-in-hand, their simultaneous development can result in the most effective coordination and implementation.

Source:
Kanjee, Anil; Acana, Sylvia. 2013. Developing the enabling context for student assessment in Uganda. Systems Approach for Better Education Results (SABER) country report ; 2013; Systems Approach for Better Education Results (SABER) student assessment working paper no. 8. Washington, DC : World Bank Group.
http://documents.worldbank.org/curated/en/2003/01/17591338/uganda-developing-enabling-context-student-assessment