Learning Metrics Partnership

A capacity support and policy strengthening initiative to develop and use common learning metrics for mathematics and reading

18 November, 2014

UNESCO United Nations Educational, Scientific and Cultural Organization

UNESCO INSTITUTE for STATISTICS

Australian Aid

Australian Government
Department of Foreign Affairs and Trade

ACER

Contents

Abbreviations

ADEA	Association for the Development of Education in Africa
EFA	Education for All
EGRA	Early Grade Reading Assessment
EGMA	Early Graded Mathematics Assessment
EMIS	Education Management Information System
GPE	Global Partnership for Education
IRT	Item Response Theory
LEG	Local Education Group
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
LMP	Learning Metrics Partnership
LMTF	Learning Metrics Task Force
PASEC	Programme d'Analyse des Systèmes Educatifs des Pays de la Confemen
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SEAMEO	South East Asian Ministers of Education Organisation
TIMSS	Trends in International Mathematics and Science Study

Poor quality education is jeopardizing the future of millions of children and youth across high-, medium- and low-income countries alike. Yet we do not know the full scale of the crisis because measurement of learning achievement is limited in many countries, and hence difficult to assess at the international level. A global data gap on learning outcomes is holding back progress on education quality.

(LMTF, 2013, p. 9-10)

Background for the Learning Metrics Partnership

Almost two thirds of all developing countries have sought to measure education quality by implementing national or participating in regional or international learning assessment initiatives (Best et al., 2013). However, these assessments vary in approach, methodology, reliability, validity and comparability. Despite the high level of participation in learning assessments, clearly defined learning metrics and intra- as well as inter-assessment comparability remain limited. This presents particular challenges for the global development goals for the post-2015 discussions and reporting requirements.

Initial work led by the Learning Measurement Task Force (LMTF) has been advanced by the UN's Open Working Group (OWG) on Sustainable Development Goals and UNESCO's Education for All Global Monitoring Report (EFA GMR). This work has reached the point where goals for the provision of quality education and for educational outcomes have now been drafted. The goals are to be supported by targets referring to learning outcomes. Learning goals and targets in the post-2015 agenda will only be meaningful if they are underpinned by empirically derived common numerical scales that accommodate results from a range of different assessments of learning outcomes. The development of common described scales allows policy makers, education practitioners and education investors to not only quantify student proficiency, but also describe it in a meaningful way. A scale provides a means to assess the emerging competencies of younger children, and to explore cognitive growth and trends over time. *A common described scale for reading and mathematics, spanning learning from early primary school to early secondary school, that is relevant and applicable to a range of developing country contexts is currently unavailable.*

The Learning Metrics Partnership (LMP) is a joint initiative of the UNESCO Institute for Statistics and the ACER Centre for Global Education Monitoring (ACER-GEM) to develop a set of nationally and internationally comparable learning metrics in mathematics and reading, and then to facilitate and support their use for monitoring purposes, in partnership with interested countries. This document outlines the LMP's three-phase program that aims to develop and validate common learning metrics for reading and mathematics, and to support countries to report the results of their assessment activities against these learning metrics. The key features of the program are fourfold:

- It accommodates results from a range of different assessments of learning outcomes.
- It yields high quality data that are nationally relevant and internationally comparable.
- It emphasises peer-to-peer capacity support and learning opportunities.
- It has a strong focus on improving data use and policy interface.

LMP Objectives and Outputs

The LMP's objective is to develop empirically derived learning metrics in mathematics and reading that will support national governments to effectively measure and monitor learning outcomes for policy purposes. The LMP does not involve the development of a new test or testing program. Rather, it supports the use of existing assessments of various kinds, and a pool of calibrated items that could be used to facilitate measurement and reporting of learning outcomes against common metrics.

The key outputs of the LMP will be:

1. common learning metrics for reading and mathematics, spanning learning from early primary school to early secondary school;
2. a set of tools and methodologies that permit the broad alignment of existing learning assessments with the common metrics; and
3. a support (capacity development) framework that enables countries to use the tools and methodologies to report results of national or other assessments against the common metrics, should they wish to do so.

What is a Learning Metric?

Any expression of learning goals should be supported by well defined indicators, which in turn draw upon accepted learning metrics and benchmarks. The process of setting and monitoring learning goals must have at its core a set of agreed learning metrics so that terms such as *foundation skills* and *acceptable* (in terms of proficiency) can be used with the knowledge that they carry a shared and accepted meaning.

For the purposes of this document the following terminology is used.

- Metric*:* This term is used to indicate a dimension of educational progression. For example, a developmental scale of reading or mathematics would be considered a *learning metric*. The metric is depicted as a line with numerical gradations that quantify how much of the measured variable (e.g. reading ability) is present. Locations along this metric can be described by numerical scores or substantively (i.e. in terms of student skills, understanding and competencies).
- Proficiency Scores: When the locations are described numerically, they are referred to as *proficiency scores*, and they quantify different performance standards for the metric.
- Proficiency Description: When the locations are described substantively, they are referred to as *proficiency descriptions.* It is not practical to develop a proficiency description for each proficiency score on the numerical scale, so proficiency descriptions are usually developed to cover particular segments of the scale. These segments are called *levels*. The *proficiency description* for a particular level can then be understood as describing the skills and proficiencies of students who attained proficiency scores that are within that particular segment of the scale. Those students would also have the skills described for lower levels.

- Benchmark: When a location is set on a scale this is referred to as a *benchmark*, which is a point on the scale against which comparisons can be made. For example, a score of 115 might be a benchmark for acceptable performance after the completion of primary schooling.
- Indicator: An *indicator*, in this context, is a quantitative expression that is used to describe the quality, the effectiveness, the equity or the trends of a particular aspect of the education system. It does so through mathematical statements concerning metrics, proficiency scores and benchmarks. For example, *the number of students at or above a specified benchmark score* could be used as an indicator of achievement in relation to that benchmark.
- Goal and target: A *goal* is often a broad aspirational statement of desired outcomes. A *target* is a specific statement of intended improvement in some particular outcome for a particular population or sub-population of interest, quantified in relation to the benchmarks, and the achievement of which can be monitored through measurements of progress on the indicators within a specified timeframe. For example, a target might be *to reduce by 30% over the next three years the proportion of Grade 3 students below the benchmark for Grade 3*; or *to increase to 90% the proportion of students completing primary school who have met or exceeded the proficiency level defined by the Grade 6 benchmark, by 2025*. Targets may vary by jurisdiction according to, for example, the current state of educational progress of children, local policy priorities, and the availability of funding to support learning interventions.

An example of a learning metric for mathematics is shown in Figure 1. The central elements of the learning metric are the numerical scale, and the descriptions of the levels of the scale in meaningful substantive terms. The various locations on this metric are proficiency scores. Given agreement on the metric, assessment tools can be developed and locations on the scale can be chosen as benchmarks, of which two have been displayed: *Grade 3 benchmark* (which may be an appropriate yardstick for some countries), and *Acceptable minimum standard for end of primary school*.

Against the learning metric in Figure 1, the learning outcomes of two countries at Grade 3 and Grade 6 are reported. For each grade for each country, a range of indicators is shown: the distribution of performance; the mean proficiency scores for all children; and the mean proficiency scores for girls, boys, urban children and rural children. A range of other indicators could also be highlighted – growth over years, differences between subgroups and so on.

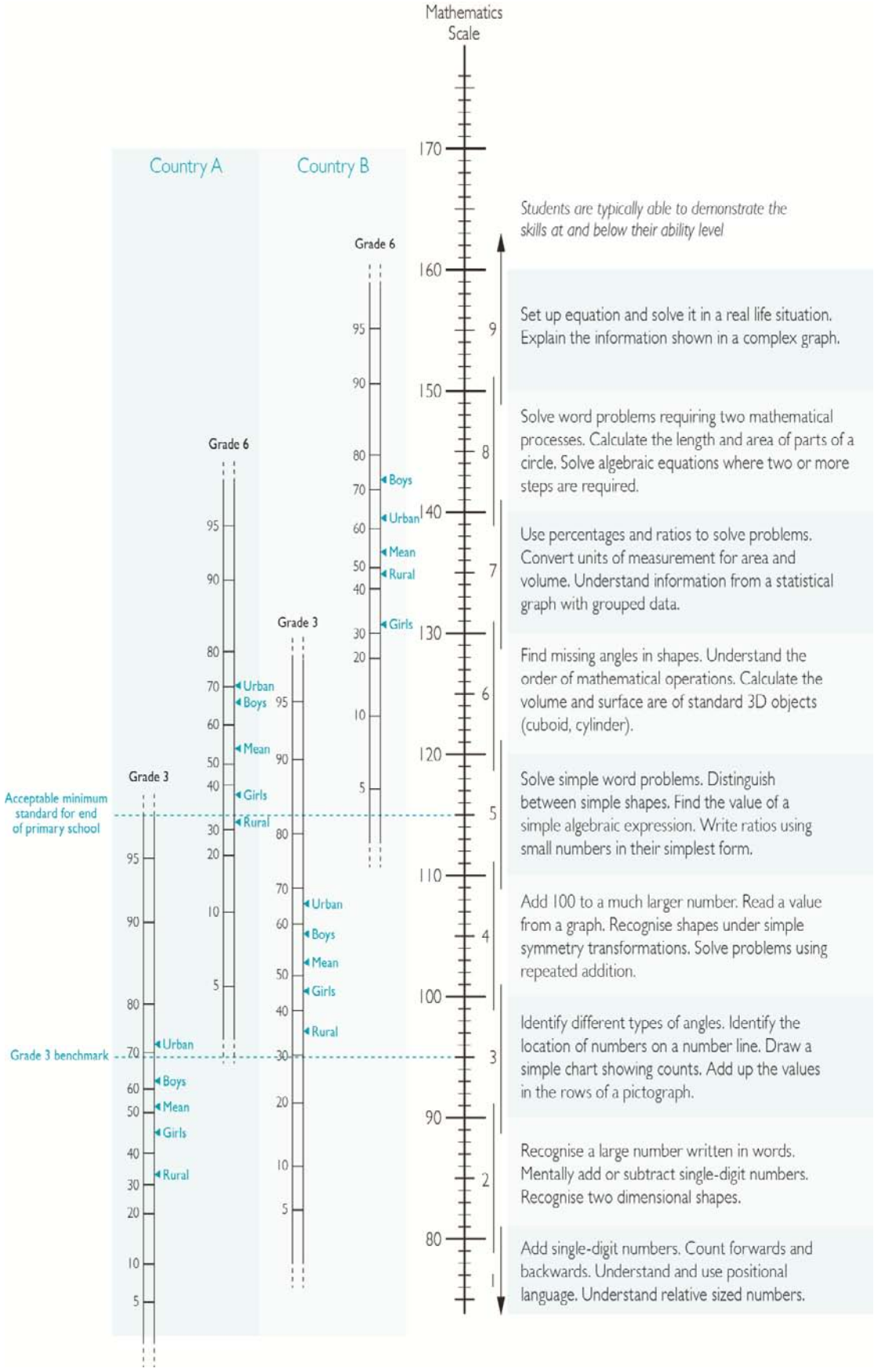

Figure 1: Example learning metric for mathematics

Description of the LMP

The LMP aims to balance two seemingly competing necessities: the necessity for common learning metrics to underpin meaningful learning goals; and the necessity for a global framework for monitoring learning outcomes that recognises and can accommodate country-specific contexts and activities.

A key element of the LMP is that it draws from existing student assessments and country-level experiences, ensuring that any learning metrics developed will be relevant for different countries' educational needs. The LMP is not proposing the development of a new assessment. Rather through a technically rigorous process of linking existing learning metrics, empirically validating, trialling, testing and reviewing in a set of pilot countries it is planned that the set of learning metrics developed will be locally relevant and reflective of varying country contexts, whilst offering comparability between contexts and over time.

Three specific deliverables will be developed through the LMP.

1. **Common Learning Metrics**

The LMP will develop, through its partnership approach, common metrics in two key domains – reading and mathematics. This will take a two pronged approach: the first will be a conceptual and the second an empirical linking exercise. The metrics will cover the range of skills and abilities tested by large-scale international and regional assessments such as PISA, PIRLS, TIMSS, SACMEQ, LLECE and PASEC, but also extend down to more foundational levels of competence that are tested by ASER, Uwezo, EGRA, EGMA.

2. **Alignment of Existing Learning Assessments with the Common Metric**

Technical work with other regional or national assessment programs will support their alignment to the common metrics. The preferred mechanism to achieve this is to form a pool of items from which a selection could be made for incorporation into existing assessments, and using those items as the basis for linking the other assessments with the common metric. Alternatively, special linking studies could be conducted to align existing assessments with the common metric.

3. **Country Level Implementation and Support Framework**

This deliverable is focussed on the application of the metrics in conjunction with in-country system strengthening in learning assessment. The LMP will focus on in-country and inter-country capacity support and development with a view to sharing technical assistance, experiences and perspectives and developing a set of tools and methodologies to systematically report results against the common metrics as part of the ongoing implementation of existing national, regional, or international assessments.

The LMP is presented as a joint initiative between the UNESCO Institute of Statistics (UIS) and the Australian Council for Educational Research. UIS is the statistical branch of the United Nations Education, Scientific and Cultural Organisation (UNESCO). The Institute produces the data and methodologies to monitor trends at national and international levels. It delivers comparative data for countries at all stages of development to provide a global perspective on education, science and technology, culture, and communication. Based in Montreal, Canada, the

UIS was established in 1999 with functional autonomy to meet the growing need for reliable and policy-relevant data. The Institute serves UN Member States, UNESCO and the UN system, as well as a range of inter-governmental and non-governmental organisations, research institutes and universities.

The Australian Council for Educational Research (ACER) is an education research organisation committed to improving the quality of learning in preschools, schools, universities and in technical and vocational education and training settings. ACER supports data collection, analysis and interpretation to support decision making related to policy formation, program planning, capacity development, monitoring and evaluating interventions to improve learning. ACER is an independent, not-for-profit organisation established in 1930 whose primary driver is improvement of learning. ACER and the Australian Department of Foreign Affairs and Trade Australian Aid program have recently established the Centre for Global Education Monitoring (ACER-GEM). This centre aims to support the monitoring of educational outcomes in a range of international contexts, holding the view that systematic and strategic collection of data on education outcomes, and factors related to those outcomes is a critical element of improving educational progress for all learners.[1]

LMP Implementation Phases and Duration

The LMP proposes to work closely with a range of education assessment specialists, development partners and Ministries of Education, to develop a more detailed plan of project activities. The LMP proposes three key phases. Project outcomes of Phase I and Phase II are expected within a **36 month** period beginning June 2014 through to June 2017.

Phase I: Drafting the Learning Metrics

The purpose of this phase is to develop a set of draft reading and mathematics learning metrics. Each metric will comprise a graduated scale and a set of descriptions of what individuals at various locations on the scale are typically able to do, illustrated by a selection of items spread along the scale. In the interest of timeliness this first phase will be undertaken without the collection of new data from students – that is it will draw upon pre-existing performance data. It is expected that the drafting of the common learning metrics will take approximately six months.

Step 1: Developing a conceptual growth framework

Conceptual growth frameworks spaning from the early stages of growth in mathematics and reading through to later stages will be outlined. These growth frameworks will be based on well established educational learning theory and informed by curriculum scope and sequence documents.

Step 2: Identifying suitable existing assessment programs

The LMP initiative does not aim to develop new test items but rather conduct a comprehensive analysis of existing items from a suitable range of assessment programs, mapping these items

[1] Further information on the ACER-GEM centre can be seen at: http://www.acer.edu.au/gem.

against the draft mathematics and reading metrics and then calibrate these items across assessments. In order to do this, the LMP will work closely with a range of assessment programs in order to jointly review these instruments.

Items from some potentially suitable candidates are already on hand or are in the public domain (for example, ASER and ASER-like instruments such as Uwezo, and the EG*A instruments). Where permission can be gained and timelines permit, instruments from programs including PASEC, SACMEQ, LLECE, PILNA, TIMSS Numeracy, PIRLS Literacy, and any others deemed relevant will also be included. In addition, there may be some national assessments (eg Afghanistan, Zimbabwe) that could provide useful information.

Step 3: Conceptual analysis of assessment items

The cognitive demand of an agreed set of items used in a variety of existing assessments will be analysed and conceptually mapped against the draft growth framework. Items will need to be obtained from as many assessments as it is feasible to include given the time constraints on the work. Assessment programs should cover learning from foundation/reception to early secondary and represent a range of item difficulties and the knowledge, skills, contexts and abilities each program attempts to measure.

Step 4: Empirical analysis of assessment items

To support the drafting of the learning metrics, data may exist from assessments that can be used to align items from each source assessment program. Some assessments using different methods of administrations, such as one-on-one oral administration, or paper-based group administration may provide comparative analysis that can be mapped against a scale using Item Response Theory (IRT) techniques. In addition, a pairwise comparison of items will be conducted to enable the different assessments to be approximately aligned. Pairwise comparison in this context refers to a process where item development specialists compare pairs of test items and judge the relative difficulties of the items in each pair. Well-established procedures (Bradley and Terry 1952; Luce 1959) will be applied to develop an estimated alignment of all available items along a single scale.

Step 5: Formulating draft proficiency descriptions

In this step information from the previous steps will inform the formulation of descriptions of growth according to the empirical difficulty of tasks used to assess elements of the conceptual framework. This step will therefore construct *separate draft learning metrics for reading and mathematics*. They will be connected to some or all of PISA, PIRLS, TIMSS, SACMEQ, LLECE and PASEC scales, but will be extend down to more foundational levels of competence.

Phase II: Validating the Metrics

The draft metrics developed during Phase I are based on the conceptual analysis of the relative difficulties of items across assessment programs, and the analysis of already existing datasets. In Phase II, the draft metrics will be validated at the country level. Data will be collected by administering combinations of items to children, which will enable the empirical determination of the relative difficulties of items across assessment programs. The LMP will adopt an item-

based approach to linking the student data[2] that will result in a pool of calibrated test items from which any country that wished to could select items and insert them into its own assessment, and so have the option of reporting its results against the common metrics.

This phase of activities will therefore involve multiple linking exercises of items from existing assessments against the draft metrics across different countries, including assessments used in Phase I and other assessments not yet used. The start-up of activities in this phase will see extensive consultation with the view to working with at least 15 countries across different continents. A clearly defined coordination mechanism will be established to facilitate strong cross-country peer support. In-country technical teams will be identified and through a process of cross-country consultation and collaboration, country-specific plans for test administration will be developed.

Phase II will have five outputs. The first will be a pool of calibrated items. The second will be an empirically-based update and validation of the draft learning metrics that were developed via conceptual alignment in Phase I. The third will be performance benchmarks set on the metrics using an empirical standard-setting exercise. The fourth will be a mapping of performance on items from the assessments used in this phase onto the common metrics. The fifth will be the establishment of a peer-to-peer capacity support coordination mechanism across multiple country locations.

The validation phase is expected to take approximately 30 months, commencing once the draft metrics have been developed in Phase I. A series of steps to implement Phase II are proposed as follows:

Step 1: Assessment programs and country participation, and coordination structures

The LMP will identify assessment programs suitable to participate in Phase II work and attempts will be made to secure their involvement. The LMP will work with existing international coordination bodies involved in the development of learning metrics including the LMTF, SACMEQ, LLECE, PASEC, SEAMEO, ADEA and others to seek country-level interest in participating in Phase II. To ensure geographical, cultural and language representation the LMP hopes to work with one to two countries each from the following nine regions:

Africa (Northern); Africa (Sub-Saharan); Africa (Eastern); Asia (Eastern); Asia (South-Eastern); Asia (Western); Oceania; Latin America and the Caribbean; Caucasus and Central Asia.

To improve the flow of information, and to take advantage of cross-country peer support and capacity exchange opportunities, a coordination framework is proposed. The framework would include a Secretariat, a high-level Steering Committee, an extended Steering Committee that

[2] There are two main approaches to equating student data: *test based* and *item-based*. The *test-based approach* is considered the most technically rigorous as assessments are administered in their complete and original test form. However, any additional country that wished to place results of its assessment program against a metric that has been validated in this manner will need to undertake a full test-based equating exercise. An alternative is an *item-based approach* where different combinations of items from a range of assessment programs are administered in different countries with the aim to establish a large bank of equated items. It is the item-based approach that is being advocated here.

includes representatives of in-country Task Teams, which could take the role of a Reference Group, and in-country technical teams. An outline of the coordination framework is presented in Figure 2.

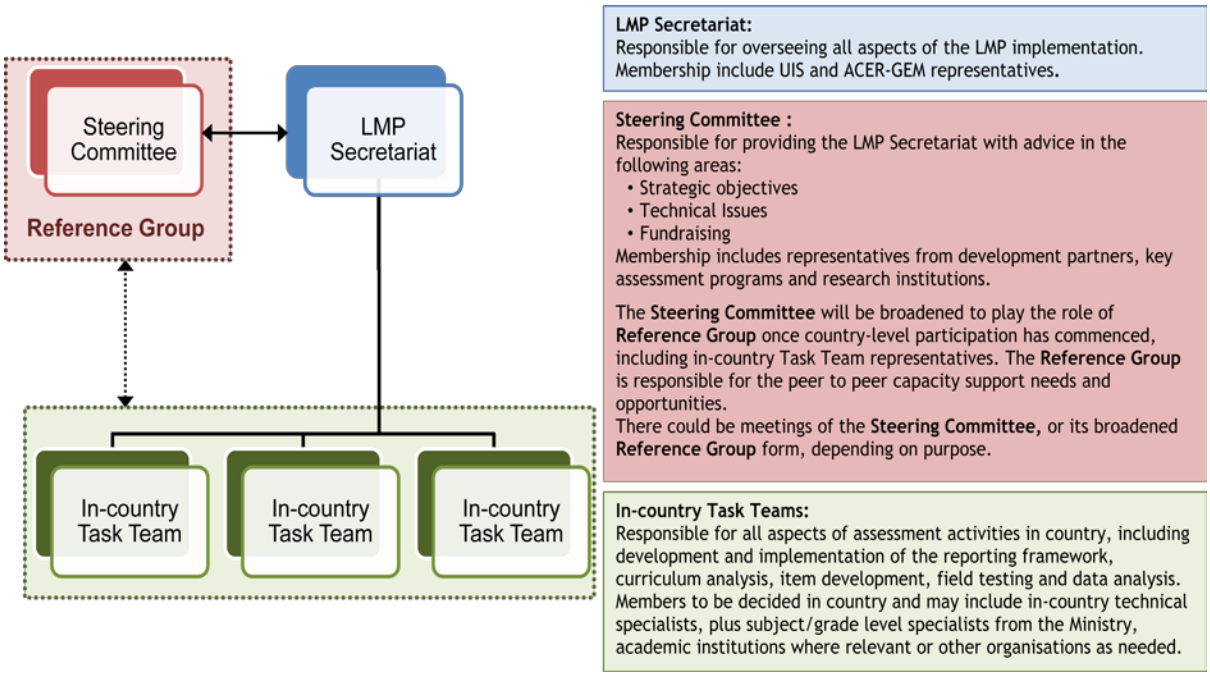

Figure 2: Proposed coordination framework for the LMP

Step 2: Selecting the items

Items selected by the experts from the assessment programs in Phase I will be again reviewed by the relevant Reference Group members to ensure there is adequate coverage of the skills, knowledge and abilities. The Reference Group will also assess what additional items should be considered for inclusion. It is expected that Phase I will not have drawn on items from all assessment programs that ideally would be represented in Phase II, because some of these assessment programs keep their items secure (e.g. SACMEQ, PASEC, LLECE). In this step it will therefore be important to obtain some of these secure items.

Step 3: Designing the tests

In this step it will be necessary to determine which combinations of item sets from different assessment programs will be administered in each participating country.

It will also be necessary to determine how many sub-populations will be assessed in each participating country. For example, which grade levels, or whether regional populations should be considered (such as when different regions use different languages).

After the mix of items to be calibrated in any one country and the sub-populations to be assessed have been determined, an appropriate technical test design for each country will be developed. The test design will give the testing time per child and the sequence of items in different test forms. It will also show how items will appear in multiple test forms to facilitate linking.

At this stage it is expected that sample sizes will be in the range of 500–1000 per population–country combination. The student sample size is not intended to be representative, but rather provide the means to empirically calibrate the relevant test items, including accommodating language coverage. The population size for the sample therefore will not be as large as for a national student assessment initiative. The expertise and knowledge residing in the Reference Group will be tapped to make the decisions needed.

Step 4: Preparing test materials

The test materials are likely to be different for each country and will depend on the items that are being administered. If a population–country combination is using items that are delivered one-on-one and orally, the test materials might comprise a test administrator's stimulus booklet, a data collection sheet on which the test administrator can record the children's answers, and an associated manual to support test administration. If a population–country combination is using items that children must answer independently, then the test materials might comprise a test booklet on which a child writes his or her answers directly, and an associated manual to support test administration.

Test materials will be developed for each country through an association between the Task Teams and the Reference Group members.

Step 5: Preparing for and undertaking data collection

In this step preparations will be made for the in-country activities. These preparations will include:

- sourcing and training test administrators
- obtaining a sampling frame and sampling children to undertake the assessment
- taking steps to identify and secure appropriate sites for test administration
- sourcing and training data entry personnel (if applicable)
- sourcing and training personnel to code student responses (if applicable).

Since each population–country combination will be completing different test forms, training for test administration and the administration itself will vary from one population–country combination to the next. It will nevertheless be important to ensure in this step that preparations are made for test administration methods that are of an agreed level of standardisation where appropriate.

Sampled children will undertake the assessments and the resulting data will be captured. Methods for data capture could include data entry into a tailored software application or scanning. Again, it may be that the methods for data capture vary across the population–country combinations.

The in-country Task Teams will lead the activities in this step and be supported by relevant test development and administration experts from the Reference Group, and other agencies where appropriate. In-country training programs will be agreed between the Task Teams and the Reference Group members prior to the start of this step.

Step 6: Analysing data and setting benchmarks

Once all data have been captured and scored, analysis will be undertaken in partnership with the in-country Task Teams with the support of relevant members of the Reference Group. Various modern psychometric techniques such as item response modelling will be employed.

This stage will also involve finalising the benchmarks. This will be an activity that requires collaboration between the Reference Group and the in-country Task Teams with relevant curriculum experts from the participating countries. In order to ensure that the benchmarks are valid for countries beyond those that participated in the linking exercises, the consultation process could be widened to include representatives from other countries that intend to make use of the metrics. Individual countries may request additional training programs by the reference group to support data analysis work.

Step 7: Mapping assessment results onto the metrics and dissemination of results

After the data have been analysed, the metrics validated and the benchmarks set, the next step will be the preparation of material that describes how countries that are involved in the assessment programs that contributed items to the linking studies can map their results onto the metrics should they wish to do so.

The Reference Group with the support of ACER and UIS will prepare this material in collaboration with the involved assessment programs. It will be the beginning of the suite of tools and methodologies that will be further developed in Phase III.

It is intended that the results relating to the development of common metrics will be disseminated as widely as possible to best inform the start up of activities related to Phase III of the program. Lessons learnt and recommendations for the implementation of student learning assessments and its related policy implications will be documented for the purpose of sharing amongst the international education development sector. The Reference Group, LMP Secretariat and other government representatives will determine the most effective dissemination strategy for the LMP results.

Phase III: Country Level Implementation

Phase III is the development of a set of tools and methodologies that permit the broad alignment of existing learning assessments and also the development of tools and methodologies to support the alignment of country-developed assessments with the common metrics.

This phase of activities will have as its major objective the development of a strategy to support country-level activities through a longer-term capacity-building partnership. The LMP recognises that every country context will have different needs from a student assessment monitoring program. For many countries, test materials and methodologies will already be well established at the country level and only slight adjustments may be needed so that reporting can be made against the common metrics. In other cases, a range of testing materials and methodologies can be available to countries who may wish to review and extend their own programs.

Establishment or strengthening an educational monitoring program, which is a central and ongoing focus of Phase III, will recognise that the most important element of any assessment

program is that it is designed so that it can inform key policy issues. The use of the tools and materials developed through the LMP will allow governments to make comparisons of data across contexts, against benchmarks, monitoring trends over time and monitoring educational growth. This approach allows users to attach real meaning to assessment outcomes, informing the next steps needed to drive improvement.

For example the inclusion of multiple grade levels in an assessment allows for information about cohort **growth between grade levels**. Information about cohort growth sheds lights on how much value is being added to students' education at different stages of their schooling, and can help education practitioners and policy makers identify the stages at which policy interventions may be required.

Additionally, an ongoing assessment program yields information about **trends over time**. This information can come in a variety of forms, including information about changes in achievement outcomes at specific grade levels or within particular sub-populations. If the program assesses multiple grade levels, then the trend information can also include details of changes in growth between grade levels over time. Trend information such as this can assist in tracking the impact of educational reforms, and guide the development of new policy.

A student assessment program must be designed to meaningfully inform policy and sector reform initiatives. In order for this to occur, it is recommended that countries undertake a policy mapping exercise prior to commencing any work on an assessment program. The aim of a policy mapping exercise is to undertake a stock take of current education policies and levels of education provision at the national, sub-national and school level. Policies related to teacher support and professional training, curriculum, school financing and school fees, provision of learning materials, hours of learning, examination systems, school quality assurance, school feeding, and school management councils, are all areas that can have an impact on learning outcomes. Whilst a comprehensive policy mapping exercise may be difficult to implement, a broad understanding of the education policy context and educational statistics is critical. Once the needs of a student assessment program become clearer so will the capacity requirements to undertake the task at hand.

Step 1: Capacity analysis

A key step in understanding different countries' strengths and program priorities as well as opportunities for peer-to-peer capacity support is an in-depth, country-specific, capacity analysis. A capacity analysis could consider areas such as:

- leadership and vision
- institutional roles and responsibilities
- staff capacity (match between staff assigned and tasks required)
- work environment (physical capacity of the workplace to service program needs)
- technical capacity and needs analysis (including more detailed capacity support plan)
- sustainability (institutional, technical, financial).

A capacity analysis can form the basis of an assessment plan that outlines all aspects of program design (technical and financial requirements), plan of activities, timelines, roles and

responsibilities, and expected outcomes. Assessment plans will build on, support and strengthen existing activities in each country.

Step 2: Capacity support for designing and implementing an assessment program

Whilst every program will be different, to ensure that assessment results are able to meaningfully inform policy there are a number of technical elements of the program that need to be considered. These can be summarised in Figure 3.

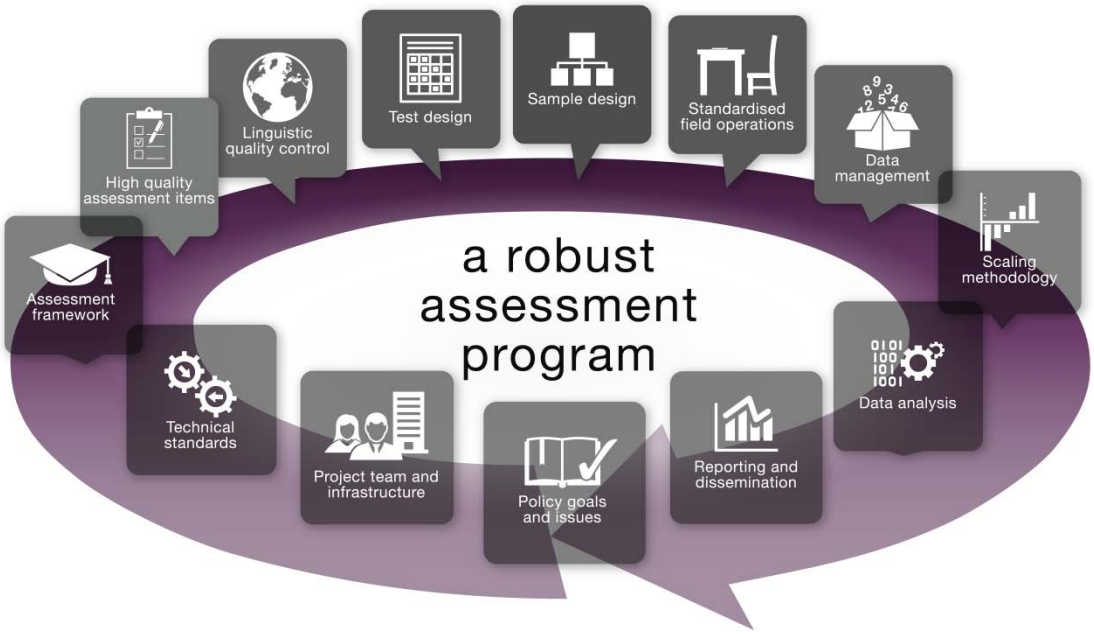

Figure 3: Elements of a robust assessment program

The LMP aims to provide the mechanisms to support individual countries in any one of, or all elements of the above mentioned areas. This request for support can be made through in-country education coordination bodies as the Local Education Group (LEG) in the case of GPE members. The LMP Reference Group members can provide support to country level education coordination bodies to help define the type and scope of support required. Direct support may take the form of tendering for large scale programs, specific short term technical assistance or longer-term tailor-made training programs. The advantage of the LMP initiative is that relevant technical expertise either through the reference group or the in-country task teams will have been mapped, which will provide opportunities for country twinning and/or peer-support initiatives.

Most assessment programs typically require one to two years to prepare. Assessment frameworks, capacity analysis, policy mapping, technical teams, test design, item development, field operations manuals, piloting, data collection and analysis, all need to be developed and completed prior to the roll-out of a full-scale assessment. Data for the first year of the assessment form the baseline for ongoing assessment. Assessment programs should aim to be integrated into national planning and monitoring frameworks of education sector plans in the same manner as (and if possible linked to) education management information systems (EMIS).

Implementing a national student assessment program therefore will take on different forms in different countries, depending on the policy requirements, the available capacity in-country, the level of financial resources and the student population size. It is recognised that every country's requirements will be different. The strength of the LMP is that it can provide tailored country level technical support to build on and strengthen existing student assessment programs, whilst allowing each country to use the products of Phases I and II to report learning assessment results against an internationally recognised set of metrics for mathematics and reading.

Proposed Budget

The budget required to implement the LMP will be dependent upon the number of countries involved and the level of existing funding in each country project.

Phase I is being funded by the Australian Department of Foreign Affairs and Trade's Australian Aid Program and ACER through ACER-GEM. Additional funding is sought to implement Phase II. Funding for Phase III is expected to be sourced from in-country government and donor-supported funding allocations. Phase III activities will vary widely by country depending on the extent of assistance required and the scope of the assessment activities. [3]

The following financial requests therefore relate to the second phase of the LMP. Approximately USD500,000 will be required for technical assistance costs related to the validation of the draft learning metrics. The linking and comparative analysis work is planned to commence from the beginning of 2015, and will entail an intensive level of time-on-task.

The proposed budget for Phase II in-country work will be dependent upon which countries wish to be involved and to what extent they request an expansion of their existing in-country activities. On average however, it is anticipated that approximately USD150,000 per country per year will be required for technical assistance, with in-country costs calculated additionally (noting that in-country costs again will vary depending on logistical requirements and existing infrastructure arrangements). Robust validation will require participation of countries from each of the nine regions listed above.

Phase III costs are more difficult to estimate and will depend upon the specifics of each country's approach to implementation. It is expected, however, that an amount of approximately USD1,000,000 over a period of three years would be required for in-country capacity development and training initiatives.

[3] Costs of technical support projects for student assessment programs can range from between USD 200,000-1,000,000 per year of support, with the majority of projects between USD 400,000–700,000 per year per grade level. Other short-term training initiatives can normally be budgeted for USD 40,000 and above.

References

Best M, Knight P, Lietz P, Lockwood C, Nugroho D, Tobin M (2013). *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries*. Final report, London: EPPI-Centre, Social Science Research Uni, Institute of Education, University of London.

Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs, I. the method of paired comparisons. *Biometrika*, 39, 324–345.

LMTF (Learning Metrics Task Force) (2013). *Toward Universal Learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington, D. C.: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.

Luce, R.D. (1959). *Individual Choice Behaviours: A Theoretical Analysis*. New York: J. Wiley.

Wagner D (2011). *Smaller, Quicker, Cheaper: Improving Learning Assessment for Developing Countries*, UNESCO/IIEP.

World Bank (2006). *World Development Report 2007: Development and the Next Generation,* Washington, DC: World Bank.