SEAMEO EXPERIENCES OF PRIMARY LEARNING METRICS

DESK REVIEW

14 November 2013
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>EALAS</td>
<td>East Asia Learning Achievement Study</td>
</tr>
<tr>
<td>EAPRO</td>
<td>UNICEF's East Asia and Pacific Regional Office</td>
</tr>
<tr>
<td>EFA</td>
<td>Education For All</td>
</tr>
<tr>
<td>EGMA</td>
<td>Early Grade Mathematics Assessment</td>
</tr>
<tr>
<td>EGRA</td>
<td>Early Grade Reading Assessment</td>
</tr>
<tr>
<td>EPR Unit</td>
<td>UNESCO Bangkok’s Education Policy and Reform unit</td>
</tr>
<tr>
<td>ESCAP</td>
<td>UN's Economic and Social Commission for Asia and the Pacific</td>
</tr>
<tr>
<td>GPE</td>
<td>Global Plan for Education</td>
</tr>
<tr>
<td>LAMP</td>
<td>Literacy Assessment Monitoring Programme</td>
</tr>
<tr>
<td>LMTF</td>
<td>Learning Metrics Task Force</td>
</tr>
<tr>
<td>MDG</td>
<td>Millennium Development Goal</td>
</tr>
<tr>
<td>PASEC</td>
<td>Programme d’Analyse des Systèmes Éducatifs de la CONFEMEN (Programme on the Analysis of Education Systems)</td>
</tr>
<tr>
<td>PIRLS</td>
<td>Progress in International Reading Literacy Study</td>
</tr>
<tr>
<td>PISA</td>
<td>Programme for International Student Assessment</td>
</tr>
<tr>
<td>READ</td>
<td>Russian Education Aid for Development</td>
</tr>
<tr>
<td>RERG</td>
<td>Regional Experts Reference Group</td>
</tr>
<tr>
<td>SACMEQ</td>
<td>Southern and Eastern Africa Consortium for Monitoring Educational Quality</td>
</tr>
<tr>
<td>SEAMEO</td>
<td>Southeast Asian Ministers of Education Organization</td>
</tr>
<tr>
<td>PLA</td>
<td>Southeast Asian Primary Learning Metric</td>
</tr>
<tr>
<td>TIMSS</td>
<td>Trends in International Mathematics and Science Study</td>
</tr>
<tr>
<td>UIS</td>
<td>UNESCO Institute for Statistics</td>
</tr>
<tr>
<td>UNESCO</td>
<td>United Nations Educational, Scientific and Cultural Organization</td>
</tr>
<tr>
<td>UNICEF</td>
<td>United Nations Children's Fund</td>
</tr>
</tbody>
</table>
Table of Contents

1. Introduction .................................................................................................................. 5
2. Rationale ....................................................................................................................... 6
3. Scope and purpose ......................................................................................................... 6
4. Limitations .................................................................................................................... 7
5. Definitions of Learning Metrics and Programs .............................................................. 9
   5.1 Classroom Assessments ......................................................................................... 9
   5.2 School surveys ....................................................................................................... 9
   5.3 Examinations ......................................................................................................... 9
   5.4 Household Surveys ............................................................................................... 11
      MICS: Multiple Indicator Cluster Survey ................................................................. 12
      Young Lives ............................................................................................................ 12
      LAMP: Literacy Assessment Monitoring Programme ............................................... 12
   5.5 Hybrid Assessments ............................................................................................. 12
      *EGRA: Early Grade Reading Assessment ................................................................... 12
      EGMA: Early Grade Mathematics Assessment ......................................................... 12
      *EALAS: East Asia Learning Achievement Study ..................................................... 13
      *Literacy Boost ....................................................................................................... 13
   5.6 Large Scale Assessments ....................................................................................... 13
   5.7 International Assessments ..................................................................................... 13
      *TIMSS: Trends in International Mathematics and Science Study ............................ 13
      *PIRLS: Progress in International Reading Literacy Study ....................................... 13
      Pre-PIRLS: Pre-Progress in International Reading Literacy Study ........................... 13
      PISA: Programme for International Student Assessment ........................................... 14
   5.8 Regional Assessments ............................................................................................ 14
      *PASEC: Programme d’Analyse des Systèmes Éducatifs de la CONFEMEN (Programme on the Analysis of Education Systems) ...................................................... 14
      SACMEQ: The Southern and Eastern Africa Consortium for Monitoring Educational Quality .... 15
      LLECE: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for Assessment of the Quality of Education) .................................... 15
   5.9 National Assessments .............................................................................................. 16
   5.10 System Strengthening Programs .......................................................................... 16
      READ: Russian Education Aid for Development ....................................................... 16
      SABER: Systems Approach for Better Education Results ......................................... 16
6. Descriptive Mapping – Primary Learning Metrics ......................................................... 17
7. Lessons learned – barriers ......................................................................................... 29
   7.1 Design and Implementation ................................................................................... 29
      Lack of in-country leadership for operational activities ............................................. 29
      Misalignment of project indicators and project inputs ............................................... 30
      Inadequate funding of assessment activities ............................................................ 30
Pressure from assessments ................................................................. 30

7.2 Policy Use .................................................................................. 30
Technical soundness ......................................................................... 30
Dissemination activities ................................................................... 31
Systems Issues ................................................................................ 31

8. Innovative approaches - facilitators ............................................. 31
8.1 Dissemination ............................................................................. 32
8.2 Soundness of program ............................................................... 32

9. Summary ..................................................................................... 33
9.1 Objectives .................................................................................. 33
9.2 Target populations ...................................................................... 34
9.3 Content ..................................................................................... 34
9.4 Logistics and frequency of assessment ...................................... 35
9.5 Facilitators of successful large-scale assessments ...................... 35
Technical soundness ......................................................................... 35
Dissemination activities ................................................................... 35
Funding .......................................................................................... 35
Strong leadership ........................................................................... 35
1. Introduction

UNESCO’s global Education For All (EFA) initiative, established in 1990 in Jomtien and reaffirmed in 2000 in Dakar, mandated that all countries should improve the quality of education, broadly defined as ‘measurable learning [outcomes achieved] by all, especially in literacy, numeracy and essential life skills’ (UNESCO, 2013a) More broadly, learning outcomes are understood to encompass both cognitive and non-cognitive outcomes, such as attitudes, skills and competencies essential for participation in the 21st century (UNESCO and UNICEF, 2012).

Globally, stakeholders in education development underscored the need for improved ways of collecting data about learning outcomes in order to improve educational quality at the local, national and international levels, (Wagner, 2011) through evidence based decision and policy making. Measuring learning outcomes at the national and international levels can help policymakers diagnose system strengths and weaknesses and appropriately design policies and direct resources, as well as help stakeholders in education development to evaluate policies and programming (LMTF, 2013). As a result, since 1990 the number of countries that have participated in national (Benavot and Tanner 2007, UNESCO 2008) and international and regional learning assessments (Kamens and McNeely 2009) has grown rapidly, especially in low and middle income countries (Best et al, 2013). Specifically, over two-thirds of low and middle income countries have participated in at least one large scale assessment of learning, from 1960 – 2008 (Kamens and Benavot 2011).

Stakeholders in global education development, such as the Global Partnership for Education (GPE) (GPE, 2011) and the Learning Metrics Task Force (LMTF – led by UNESCO UIS and the Brookings Institute) emphasise the continued importance of educational quality and learning outcomes for the post-2015 era, as many school aged children are not meeting minimum standards in literacy and numeracy (LMTF, 2013). In addition, the limited evidence about learning outcomes in the Asia Pacific region shows ‘mixed and limited progress’ for improving educational quality in the region (UNESCO and UNICEF, 2012).

Despite the mixed evidence for improving educational quality, many countries in the Asia region have rich experiences with learning metrics, and have adopted systematic approaches, or multilevel approaches to monitoring learning outcomes at the local, national and international levels (Ho, 2012). Learning metrics serve different purposes depending upon the differing information needs of stakeholders and contexts.

The following simplified framework (adapted primarily from Wagner et al – Kanjee, 2012 and Wagner 2011) helps to illustrate the components of a multilevel approach to assessment, shown below in Figure 1. A comprehensive assessment system may include: classroom assessments, school surveys, examinations, household surveys, hybrid assessments, large scale assessments (international, regional and national) as well as system strengthening activities at the institutional level. According to Kanjee (2012), these assessment components are interrelated and are implemented at different levels of education (early childhood, primary and post-primary). The assessment components sit within cultural, economic, political and social frameworks, which are reflected by national policies, stakeholders, institutions and processes (ibid). Examples of well known learning metrics have been provided to help illustrate the included assessment types, but do not intend to serve as a comprehensive documentation of all assessment programs.
2. Rationale

There has been large variation between countries with experiences with learning metrics (Wagner et al. – Kanjee, 2012) for evidence based decision and policymaking. Stakeholders agree that countries should have multilevel approaches to assessment, which work in tandem to provide comprehensive information about student learning and educational quality. Yet, the current focus of education development is on learning metrics at the national and international levels (GPE, 2011; LMTF, 2013). Global learning metrics can help to monitor progress towards post-2015 goals and improving educational quality to improve programming and planning at the international and national levels, as well as increase transparency, accountability and knowledge sharing among countries and stakeholders. In addition, assessing learning outcomes in primary education can better meet the needs of students and education systems in low and middle income countries, as many children leave school at the end of primary education or secondary education (Wagner et al. – Gove, 2012). Learning metrics for primary education can provide more timely data to stakeholders and policymakers for effective reforms and interventions.

Towards this aim, the Southeast Asian Ministers of Education Organization (SEAMEO) has established the Southeast Asian Primary Learning Metric Initiative (SEA-PLM) to develop common primary education learning metrics for SEA countries, as less is known about SEA countries’ experiences with these tools (SEA-PLM Concept Note, 2013). Furthermore, less is known about the ways that these assessments are used for improving educational quality, as results from a systematic review suggest that large scale assessments impacted upon policymaking slightly less in the Asia region than in other regions (Best et al. 2013).

A working group, including the EPR Unit, UNESCO Bangkok (NEQMAP Secretariat), UIS Bangkok, ESCAP Statistics Unit, SEAMEO Secretariat, SEAMEO Innotech and UNICEF EAPRO, will help to inform the SEA-PLM Initiative. The current desk review was prepared by ACER, in consultation with UNESCO Bangkok and SEAMEO Innotech, in preparation for the Regional Experts Reference Group (RERG) Seminar, to be held in Quezon City, Philippines September 24th – 26th, 2013.

3. Scope and purpose

The purpose of this desk review is to provide a high level overview of large scale assessments in at the primary school level in SEAMEO countries with a view to informing the aims and objectives, content areas and target populations of a possible regional assessment programme.

To this end, this review will:

1) descriptively map SEAMEO member country participation with learning metrics in primary education by describing general features of learning metrics and country participation, specifically with hybrid and large scale assessments (international, regional and national); and

2) broadly examine innovative approaches and lessons learned with learning metrics for implementation, capacity building and policymaking in SEA countries, and in low and middle income countries in general.
The desk review reports and synthesises information from publicly available literature: published agency reports, journal articles, and research studies. It also includes grey literature such as: working papers, draft reports from agencies, power point presentations, and information from online databases from aid agencies.

Given the focus on SEAMEO member countries, the geographical scope of this review as follows:

Brunei Darussalam
Cambodia
Indonesia
Lao PDR
Malaysia
Philippines
Republic of the Union of Myanmar
Singapore
Socialist Republic of Vietnam
Thailand
Timor-Leste

4. Limitations
The descriptive mapping in point 1 does not aim to include SEA countries’ experiences with all assessment types, as SEAMEO Innotech undertook a comprehensive review of assessment systems in SEA countries, focused on classroom assessments, school surveys, and examinations (internal and external), as well as national assessment policies and frameworks (SEAMEO Innotech, Draft).

Additionally, there is scope to further undertake a systematic survey of SEA countries’ experiences with national and sub-national assessments and learning metrics. The current review includes experiences with national assessments that are available in the public domain, primarily compiled from UNESCO’s Global Monitoring Report 2008 and UNESCO Education and Policy Reform (EPR) Unit’s working document 1 on learning assessments and policy use (2013).

This desk review is not meant to be exhaustive but to provide a condensed snapshot of large scale assessment activity in the region to date. Case studies to provide further details regarding obstacles and risks of large scale assessments in SEAMEO countries are beyond the scope of this review.
Figure 1  Simplified framework of multilevel approaches to assessment

Cultural ↔ Economic ↔ Political ↔ Social

Stakeholders ↔ Policies ↔ National Curriculum ↔ Institutions and Processes

Early childhood → Primary → Secondary → Tertiary → Adulthood

Classroom Assessments  School surveys  Examinations  Household Surveys  Hybrid Assessments  Large Scale Assessments  System Strengthening

MICS  EGRA/EGMA  International TIMSS, PIRLS, PISA  SABER

LAMP  EALAS  Regional SACMEQ, PASEC, LLECE  READ

Young Lives  National

Sources: Adapted from Wagner et al – Kanjee, 2012; Wagner 2011; LMTF, 2012: Clarke, 2011
5. Definitions of Learning Metrics and Programs

Learning metrics can serve different purposes and be used in different ways, depending upon the information needs of stakeholders and contexts. For example, learning metrics can aim to ensure quality and/or equity of the education system, to serve as an accountability tool, or to monitor and evaluate learning progress (Best et al, 2013). Learning metrics can be used to create awareness of issues among stakeholders, to formulate policy, to implement policy or to monitor policy (ibid).

This section will broadly define the purposes, methodologies and uses of different learning metrics outlined in Figure 1. At the same time, this section will provide brief descriptions of specific primary learning metrics that are included and excluded from this review because, for instance, they are undertaken with age groups that are beyond primary schooling either in or outside school (e.g. PISA, LAMP).

This will lay the groundwork for the categories used in the descriptive mapping (* denotes inclusion in mapping) of the primary learning metrics in which SEAMEO countries have been involved to date that is shown in Table 2.

5.1 Classroom Assessments

Classroom assessments provide immediate information about individual learners to help improve teaching and learning processes in the classroom (Clarke, 2011; SEAMO Innotech, Draft). In addition, classroom assessments can be used to report to parents, to monitor student learning against curricular and learning standards, as well as to identify individual student learning needs to more appropriately allocate resources at the classroom and school levels (Ho, 2012).

5.2 School surveys

School surveys, or evaluations, assess the effectiveness and efficiency of individual schools, by using results from classroom assessments, examinations and national surveys (Wagner et al – Kanjee, 2012). School surveys are not assessments, or learning metrics themselves, but an activity undertaken to evaluate school performance (ibid), and ensure accountability for improving student learning and teaching practice and control the education quality of school.

5.3 Examinations

Examinations may also be referred to as ‘public’, ‘external’ or ‘end of cycle’ (Clarke, 2011) and are used to provide information about the learning outcomes of individual students. They are often high-stakes, and are used to certify achievement (end of primary or secondary school leaving certification), or to select students for further study (Postlethwaite & Kellaghan, 2008; Clarke, 2011). Therefore examinations are distinct from low-stakes national assessments, which report aggregate performance of students and are an indicator of educational quality of education systems (Kellaghan & Greaney, 2001).

However, at the country level, sometimes the distinction between examinations and national assessments is not clear (Ho, 2012) as some countries use examinations assessment learning outcome of individual student, or evaluate school performance, or to provide policymakers with information about student learning (ibid). There are different technical
requirements for the development and use of examinations and national assessments (Postlethwaite & Kellaghan, 2008). Therefore, caution should be exercised in using examinations for uses other than reporting on individual student performance (Clarke, 2011). Table 1 briefly illustrates the examination systems in Brunei Darussalam and Cambodia.

Table 1  Example of examination/ assessment systems in Brunei Darussalam and Cambodia

<table>
<thead>
<tr>
<th>Country</th>
<th>Examination</th>
<th>Purpose</th>
<th>Participants and Users</th>
<th>Examination Body</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunei Darussalam</td>
<td>Primary Certificate of Examination (PCE)</td>
<td>Entry to LSE or confirmation that students have completed primary level</td>
<td>Year 6 students</td>
<td>Department of Examination, MOE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Also used as basis for selection to special secondary schools (e.g. Science School)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower Secondary (Penilaian Menengah Bawah)</td>
<td>Entry to USE or confirmation students complete lower secondary school</td>
<td>Year 9 students</td>
<td></td>
</tr>
<tr>
<td></td>
<td>General Certificate of Education (BCGCE O Level)</td>
<td>Entry to Year 12 of USE</td>
<td>Year 11 students</td>
<td></td>
</tr>
<tr>
<td></td>
<td>General Certificate of Education (BCGCE A Level)</td>
<td>Entry to Higher Education or confirmation that students have completed upper secondary education</td>
<td>Year 13 students</td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td>National Examination for LSE Certificate</td>
<td>Completion of LSE and entry to USE</td>
<td>Year 9 students</td>
<td>Examination Office of the General</td>
</tr>
<tr>
<td>Country</td>
<td>Examination</td>
<td>Purpose</td>
<td>Participants and Users</td>
<td>Examination Body</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Brunei Darussalam</td>
<td>Primary School Assessment (PSR)</td>
<td>Entry to LSE / Also used as basis for selection to special secondary schools (e.g. Science School)</td>
<td>Year 6 students</td>
<td>Department of Examinations, MOE</td>
</tr>
<tr>
<td></td>
<td>Student Progress Assessments (SPA); (SPA=SPE + SBA) (SBA = School based assessment; SPE = Student Progress Examinations)</td>
<td>Entry to USE; as basis for selection to different education programmes</td>
<td>Year 8 students</td>
<td>Department of Examinations, collaboration with Department of Schools, MOE</td>
</tr>
<tr>
<td></td>
<td>General Certificate of Education ‘O’ Level (B-C GCE ‘O’) / IGCSE</td>
<td>Entry to Pre-University (Sixth Form) / Employment / Technical/higher institutions</td>
<td>Year 10 / 11 students</td>
<td>Department of Examination, MOE</td>
</tr>
<tr>
<td></td>
<td>General Certificate of Education ‘A’ Level (B-C GCE A Level)</td>
<td>Entry to Higher Education / Employment / Technical Education</td>
<td>Upper sixth students</td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td>National Examination for LSE Certificate</td>
<td>Completion of LSE and entry to USE</td>
<td>Year 9 students</td>
<td>Examination Office of the General Secondary Education Department, MOEYS</td>
</tr>
<tr>
<td></td>
<td>National Examination for USE Certificate</td>
<td>Completion of USE and entry to Higher Education</td>
<td>Year 12 students</td>
<td></td>
</tr>
</tbody>
</table>


5.4 Household Surveys
Household surveys gather specified information on target populations within countries or regions, and may be used for local and national stakeholder needs, international comparability or program assessment (Wagner, 2011). Household surveys are often used to gather information on literacy, numeracy and general skills of populations that are not in school, such as adults, young children, or out of school children. However, learning outcomes are most often assessed in schools and rarely with household surveys. Hybrid Assessments attempt to target sampling to capture information on populations of interest, such as out of school children (ibid).

**MICS: Multiple Indicator Cluster Survey**
MICS is a household survey of early childhood development that examines literacy, numeracy, physical and social-emotional development, for children aged 0 to 5. MICS has been used in 55 countries with support from UNICEF during: 1995, 2000, 2005, 2009-2011. (LMTF, 2013)

**Young Lives**
Young Lives is a household survey that examines literacy, numeracy and social-emotional outcomes for children aged 4 to 17. It has been implemented in 4 countries (including Vietnam), by a consortium of research organisations led by the University of Oxford. (LMTF, 2013).

**LAMP: Literacy Assessment Monitoring Programme**
LAMP is a household survey that examines literacy outcomes for adults aged 16 to 64. It was developed by the UNESCO Institute for Statistics (UIS), and is implemented by national governments. (LMTF, 2013). As LAMP's target population is beyond the age of primary schooling it has not been included in the descriptive mapping.

**5.5 Hybrid Assessments**

Hybrid Assessments, a term used by Wagner (2011), combine methodologies from household surveys and large scale assessments to address the contextual needs of low and middle income countries. Hybrid assessments examine learning in local contexts, rather than international comparability and use efficient and cost-effective administration and analyses. Learning metrics, or tools, produced for hybrid assessments can be adapted to be used in other contexts and for different levels of the education system (e.g. national or local) or for different purposes (e.g. program evaluation) (ibid).

***EGRA: Early Grade Reading Assessment**
An early years literacy assessment, administered orally to children in grades 1 to 4. EGRA examines basic literacy skills, necessary for literacy development: letter recognition, word recognition, sentence and paragraph structure and listening comprehension. EGRA has been developed and supported by RTI International, and USAID. EGRA has been implemented in approximately 44 countries since 2008. (LMTF, 2013; FTI, 2009).

**EGMA: Early Grade Mathematics Assessment**
An early years numeracy assessment, administered orally to children in grades 1 to 4. EGMA examines basic numeracy skills. EGMA has been developed and supported by RTI International, and USAID. EGRA has been implemented in approximately 11 countries since 2011. (LMTF, 2013; FTI, 2009). As no SEAMEO countries have implemented EGMA, it has not been included in descriptive mapping.
**EALAS: East Asia Learning Achievement Study**
EALAS is a series of learning achievement studies conducted at the national and sub-national levels in East Asian countries undertaken from 2004 to 2006. Studies were conducted at either grades 3 and 5, or grades 4 and 6 and assessed students in national curricular domains (e.g. mathematics, language, science, life skills) and studies also collected contextual information from students and schools. EALAS was supported by UNICEF, aimed to serve as national capacity building projects in assessment. (UNICEF, 2007).

**Literacy Boost**
Literacy Boost is an early grade reading assessment from grades 1 to 4, as well as literacy intervention that provides teacher professional development as well as community resources to support improved literacy outcomes for young children. Literacy Boost was developed by Save the Children in 2009, and is implemented at the national and sub-national levels. In some cases, Literacy Boost may use EGRA as a baseline assessment, or adapt EGRA, such as not timing fluency of reading. (LMTF, 2013).

### 5.6 Large Scale Assessments
Large scale assessments are used for monitoring purposes and to inform policymaking by providing information for the overall education system, such as performance levels and contextual information related to learning achievement (Clarke, 2011). Furthermore, large scale assessments aim to enable comparisons over time (Best et al, 2013), through their rigorous design, content, administration and scoring (deLandshere 1997). Large scale assessments can be broadly classified as international assessments, regional assessments and national assessments.

### 5.7 International Assessments
International assessments are conducted in various languages and regions throughout the world (Clarke, 2011) and were designed to explore cross-national variation in student learning, education systems and processes (Keeves, 1995; Best et al, 2013). Lockheed (2012) characterises international assessments by those that: involve multiple countries; use standardised instruments, implementation and analyses; sample large student populations that are comparable across participating countries (Wagner et al – Lockheed, 2012).

**TIMSS: Trends in International Mathematics and Science Study**
TIMSS is a mathematics and science assessment that monitors trends in student achievement, based on the curricula of participating countries and that is administered to students in grades 4 and 8. TIMSS is supported by the International Association for the Evaluation of Educational Achievement (IEA) and has been administered in four year cycles since 1999, with 63 countries having participated in the 2011 cycle (FTI, 2009).

**PIRLS: Progress in International Reading Literacy Study**
PIRLS is a reading literacy assessment that monitors trends in student achievement and is administered to students in grade 4. PIRLS is supported by the IEA, and has been implemented in four year cycles since 2001, with 49 countries having participated in 2011 (FTI, 2009).

**Pre-PIRLS: Pre-Progress in International Reading Literacy Study**
Pre-PIRLS is a basic reading comprehension assessment that caters to the needs of low and middle income countries, for students in grades 4 to 6. Pre-PIRLS is supported by the IEA, and was first administered in 2011 in three countries, and will be administered in 2016. As
no SEAMEO countries have participated in Pre-PIRLS, it has not been included for descriptive mapping. (LMTF, 2013).

**PISA: Programme for International Student Assessment**
PISA is an assessment that monitors trends in student literacy, mathematics and science achievement, based on literacies or skills and competencies in these areas, and is administered to 15 year old students. PISA is supported by the Organisation for Economic Co-operation and Development (OECD), and has been administered in three year cycles since 2000, with 66 countries and economies having participated in 2012. (LMTF, 2013). As PISA focuses on learning outcomes in secondary education, not primary education, it has not been included for descriptive mapping.

5.8 Regional Assessments

Regional assessments sample students and schools in regions that share similar historical, linguistic, cultural and economic conditions to explore cross-national variation in student achievement with ‘like’ countries (Best et al, 2013). Regional assessments intend for data and policy implications to be more relevant and therefore readily shared across participating countries.

*PASEC: Programme d’Analyse des Systèmes Éducatifs de la CONFEMEN (Programme on the Analysis of Education Systems)*
PASEC is a regional assessment that is primarily administered in francophone countries in Africa, or former French colonies. It aims at:
- informing its members of the evolution of education systems and ongoing reforms;
- providing key inputs for a reflection on topics of common interest in sight of collaborative action;
- propelling cooperation between Ministers and experts in order to elaborate common positions and propose recommendations to support regional and international policies in education” (http://www.confemen.org/wp-content/uploads/2012/02/Nouvelle_Vision_Pasec_version_anglaise.pdf, last accessed 11 November 2013)

Tests are administered to students in grades 2 and 5 in French, mathematics, health and other domains, based on common curricula of the participating countries. PASEC is supported by CONFEMEN, an organisation of francophone ministries created in 1991, with a permanent secretariat in Dakar, Senegal. (LMTF, 2013). PASEC assessments are implemented at the country level, and to date, 22 national projects have been undertaken. Until recently unique feature of PASEC was that it administered pre and post assessments to students within one academic year, in order to evaluate teaching, learning and policy.

Currently, however, PASEC is undergoing major changes, aimed at strengthening the measurement of the learning outcomes to enable appropriate comparisons across participating countries and over time. Thus, PASEC intends to measure student performance on a common scale using Item Response Methodology (IRT) that will enable the accurate performance estimates at the country level and for sub-group of interest within countries, depending on the sampling design within countries. The changed PASEC will be implemented at three levels of that are considered important points of the education system:
- Beginning of schooling (2nd grade of primary school),
- End of primary cycle (6th grade of primary school),
- End of basic education cycle (end of lower secondary school) – as of 2016.

SACMEQ: The Southern and Eastern Africa Consortium for Monitoring Educational Quality
SACMEQ's mission has been specified by the ministers of its member countries as follows:

"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". (http://www.sacmeq.org/mission, last accessed 11 November 2013)

SACMEQ is a regional assessment that is administered to students in grade 6 in Anglophone African countries, based on common curricula of the participating countries. SACMEQ was established in 1995, with support from UNESCO IIEP and the government of the Netherlands. To date, three cycles of SACMEQ have been administered across 15 participating countries. A unique feature of SACMEQ is that in some national assessments, teachers are also administered assessments to evaluate their curricular knowledge and skills. SACMEQ aims to serves to monitor educational quality in the region, and to serves to develop institutional capacity in assessment use and dialogue between researchers, policymakers and practitioners.

LLECE: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for Assessment of the Quality of Education)

LLECE and its objectives are defined as follows:

"The Latin American Laboratory for Assessment of the Quality of Education (LLECE) is the network of quality assessment systems for education in Latin America. It is coordinated by UNESCO’s Regional Bureau for Education in Latin America and the Caribbean with headquarters in Santiago, Chile. Its objectives are:

• producing information about students’ learning achievements and analysing associated-factors that explain this progress;
• supporting and advising the measurement and assessment Units of the different countries; and
LLECE is a regional assessment of students in grade 6, in Latin American and Caribbean countries. LLECE is managed by UNESCO’s Regional Bureau for Education in Latin America and the Caribbean in Santiago, Chile. Two cycles of the assessment have been administered, one in 1997 and the other in 2006. Currently the third cycle of LLECE is in progress. Initially the assessment was administered to students in grades 3 and 4, and now is administered to students in grades 3 and 6 in five year cycles. (LMTF, 2013).

5.9 National Assessments

National assessments measure and monitor learning outcomes of a whole education system, or defined part of an education system, and do not assess individual student performance (Kellaghan & Greaney, 2001), like examinations. National assessments most often assess similar curricular domains: literacy, mathematics and science, and could also include civics, art, and social science among other domains (ibid). National assessments most often are administered in primary education, and to a lesser extent, in secondary education (ibid) and can use census or representative sampling. There is cross-national variation in the frequency of national assessments, with some countries conducting a national assessment every year, and other countries conducting national assessments less frequently. Additionally, at the within country level, more established national assessment systems may be implemented regularly, while countries with new assessment systems may implement national assessments irregularly.

As national assessments are undertaken by ministries of education, there is a lack of publicly available literature about participation in these activities, describing the design and implementation of such assessments in detail, in comparison to other learning metrics. There is a greater lack of publicly available literature which discusses how evidence from national assessments is used for decision making at local and national levels. Therefore, as outlined in the Limitations section of this review, the descriptive mapping will include publicly available information about SEAMEO countries’ experiences with national assessment programs, while acknowledging that this mapping is incomplete and would benefit from further research.

5.10 System Strengthening Programs

System strengthening programs are reviews and evaluations of the institutional capacity of education systems as regards multilevel approaches to assessment. System strengthening focuses on effective policy conditions and processes for the improved design, implementation and most importantly, use of learning metrics for supporting improved educational quality.

**READ: Russian Education Aid for Development**

READ is a policy initiative led by the Government of Russia and the World Bank to broadly support improved educational quality in EFA-Fast Track Initiative countries. The initiative aims to strengthen countries’ capacity to design, implement and use student assessments. Since 2008, 8 countries have participated in READ, including Vietnam, which received support to participate in PISA 2012. (FTI, 2009)

**SABER: Systems Approach for Better Education Results**

SABER is an initiative of the World Bank that aims to provide rigorous, policy analysis and evaluation of relevant policy domains for improving educational quality in low and middle
income countries. SABER uses a systems approach to policy evaluation and focuses on policy implementation, or ‘actionable information’ to provide relevant information for policymakers. SABER has included ‘Assessment’ as a policy domain, and has produced a framework to evaluate a country’s multilevel approach to assessment. SABER aims to produce transparent, open and comparative indicators about policy choices for Assessment, or learning metrics. To date, 16 countries, including Vietnam, have participating or are undergoing a SABER review in Assessment.

6. Descriptive Mapping – Primary Learning Metrics

This section descriptively maps SEAMEO member countries’ experiences with primary learning metrics by documenting and broadly describing: the learning metric and type, purpose, assessment domains, year levels, year(s) participated, sampling, notable features and agencies/stakeholders involved. The collated information is summarised in Table 2.

Participation is mapped for hybrid and large scale assessments, which include international, regional and national/sub-national assessments. As most international and regional assessments are standardised across countries (e.g. TIMSS, PIRLS), information that has been outlined in the previous section was not repeated in Table 2. More information for hybrid and national assessments has been included, as these learning metrics can vary by country. As a result, information for hybrid assessments, particularly EGRA, have been disaggregated at the country level and presented as separate rows, as experiences vary by each cycle. Participation in more standardised assessments (e.g. TIMSS, PIRLS), have been aggregated in one row per country, and the year(s) of participation listed. Where information is unknown, the cell has been left blank.

Examining SEAMEO country experiences with primary learning metrics, it is evident that there is a range of experience, with some countries having less experience (e.g. Brunei Darussalam, Myanmar) and some countries exhibiting a breadth of experience across different types of learning metrics (e.g. Philippines, Vietnam). Participation in learning assessments has grown rapidly among low and middle income countries since 1990; experience with learning metrics at the primary level has occurred recently for many SEAMEO countries within the past decade. This finding is also reflected in a review of World Bank support for assessment activities from 1998 – 2009. Liberman and Clarke (2012) found that 12 education projects, focused on assessment, were supported across 8 countries in East Asia and the Pacific. These projects aimed to support the development of primary education assessment systems through capacity development at the institutional, policy and resource levels. However, some SEAMEO countries have a longer history with primary learning metrics, starting in the 1990s (Indonesia, Malaysia) and even 1980s (Singapore).

SEAMEO primary learning metrics mostly measure learning outcomes in literacy, mathematics, and to a lesser extent, science. This finding also reflects the domains commonly assessed across large scale assessments, as globally, literacy, mathematics and science constitute the main curricular areas taught in schools (Best et al, 2013). In a few cases, other domains assessed in primary learning metrics include social science, ICT and foreign language.

Examining the frequency of assessment types at the country level (multiple entries for EGRA within a country are counted only as 1 hybrid assessment), SEAMEO countries have more experience with hybrid assessments, then national assessments, international assessments
and regional assessments. Experiences with hybrid assessments are closely shared among EGRA, EALAS and Literacy Boost assessments. Benefits of hybrid assessments include the possibility for greater transparency in monitoring and evaluating learning outcomes, as hybrid assessments can be adapted across different levels of education (e.g. national, local program evaluation) (Wagner, 2011). Hybrid assessments are more easily adaptable to meet the needs of the local context (e.g. linguistic, out of school children, limited funding), yet, they lose comparability across education systems and administrations when adapted and administered to single populations of interest.
<table>
<thead>
<tr>
<th>Country</th>
<th>Learning Metric</th>
<th>Type</th>
<th>Purpose</th>
<th>Domains</th>
<th>Year Levels</th>
<th>Year(s)</th>
<th>Sampling</th>
<th>Features</th>
<th>Agencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunei Darussalam</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>To inform teacher practice and instruction</td>
<td>Literacy</td>
<td>Grades 1–6</td>
<td>2010</td>
<td>National sample: 40 schools, approximately 2400 students</td>
<td>Language of administration: Khmer; Writing Tools used for program evaluation i.e. multiple equated forms used to track progress over time</td>
<td>Ministry of Education. EFA - Fast Track Initiative</td>
</tr>
<tr>
<td>Cambodia</td>
<td>PASEC</td>
<td>Regional</td>
<td>Diagnostic evaluation</td>
<td>Literacy, Mathematics</td>
<td>Grades 2, 3, 5, 6</td>
<td>2012, 2013</td>
<td>Representative sample of public schools</td>
<td></td>
<td>CONFEMEN</td>
</tr>
<tr>
<td></td>
<td>Learning Assessment System</td>
<td>National</td>
<td></td>
<td>Literacy (Khmer), Mathematics</td>
<td>Grades 3, 6</td>
<td>2006 - 2008</td>
<td></td>
<td></td>
<td>Ministry of Education, World Bank</td>
</tr>
<tr>
<td></td>
<td>National Study of Student</td>
<td>National</td>
<td>To determine Level of competency or proficiency in English Language and Mathematics</td>
<td>Literacy, Mathematics</td>
<td>Year 4, 6, 8</td>
<td>2008</td>
<td>All students in Gov. Schools</td>
<td></td>
<td>Department of Planning, Development and Research, Ministry of Education.</td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------</td>
<td>----------</td>
<td>---------------------------------------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>---------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td><strong>Indonesia</strong></td>
<td>Literacy Boost</td>
<td>Hybrid</td>
<td>Baseline assessment of early years reading</td>
<td>Literacy</td>
<td>Grade 2</td>
<td>2012</td>
<td>673 student from 35 schools in one province</td>
<td></td>
<td>Save the Children</td>
</tr>
<tr>
<td></td>
<td>PIRLS</td>
<td>International</td>
<td></td>
<td>Literacy</td>
<td>Grade 4</td>
<td>2006, 2011</td>
<td></td>
<td></td>
<td>IEA</td>
</tr>
<tr>
<td></td>
<td>Assessment of Students Learning Achievement</td>
<td>National</td>
<td></td>
<td>Literacy (Indonesian), Mathematics, English</td>
<td>Grade 3</td>
<td>2005 - yearly</td>
<td></td>
<td></td>
<td>Educational National Standard Board</td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Lao PDR</td>
<td>EGRA</td>
<td>Hybrid</td>
<td></td>
<td>Literacy</td>
<td></td>
<td>2013</td>
<td>3 regions, 3200 students</td>
<td>Tools used as national or system-level diagnostic i.e. representative sample was drawn, and inter-rater reliability was investigated</td>
<td>World Bank</td>
</tr>
<tr>
<td></td>
<td>PASEC</td>
<td>Regional</td>
<td>Diagnostic evaluation</td>
<td>Literacy, Mathematics</td>
<td>Grades 2, 4</td>
<td>2012-2013</td>
<td>Nationally representative sample</td>
<td></td>
<td>CONFEMEN</td>
</tr>
<tr>
<td></td>
<td>Assessment of Student Learning Outcomes</td>
<td>National</td>
<td>Monitor educational quality</td>
<td>Literacy, Mathematics</td>
<td>Grade 5</td>
<td>2006</td>
<td>Representative sample; 7450 students</td>
<td></td>
<td>Ministry of Education, I Research Inst. for Educational Science</td>
</tr>
<tr>
<td></td>
<td>Primary School Achievement Test</td>
<td>National</td>
<td></td>
<td>Malay, English, Mathematics, Science, Chinese, Tamil</td>
<td>Grade 6</td>
<td>1987 - yearly</td>
<td></td>
<td></td>
<td>Ministry of Education; Malaysian Examination Syndicate</td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Myanmar</td>
<td>EALAS, also known as Learning Achievement Study</td>
<td>Hybrid</td>
<td>Capacity building project; assess learning achievement in curricular domains</td>
<td>Mathematics, Literacy, Science, Social Science</td>
<td>Grades 3, 5</td>
<td>2005</td>
<td>20 schools</td>
<td>Contextual questionnaires for schools</td>
<td>UNICEF</td>
</tr>
<tr>
<td>Philippines</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>Baseline for evaluating learning outcomes for sponsorship-funded programming in Metro Manila and South Central Mindanao program sites; and to inform Literacy Boost programming in selected communities</td>
<td>Literacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>Baseline for a Whole School Reading Program; Inform training for teachers; inform instruction</td>
<td>Literacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------</td>
<td>-------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Philippines</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>Project deliverable - tool used for a research study on program efficacy</td>
<td>Literacy</td>
<td></td>
<td>2012–2013</td>
<td>810 students for pre- and post-test</td>
<td>Languages of administration: English. Tools used for program evaluation i.e. multiple equated forms used to track progress over time</td>
<td>USAID, EDC,</td>
</tr>
<tr>
<td>Philippines</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>National baseline prior to implementation of MTE–MLE curriculum</td>
<td>Literacy</td>
<td></td>
<td>2013–ongoing</td>
<td>2,810 students, 40 schools nationwide</td>
<td>Languages of administration: English, Filipino (Tagalog), Ilokano. Subtasks tested: to be determined. Tools used as national or system-level diagnostic i.e. representative sample was drawn, and inter-rater reliability was investigated</td>
<td>USAID, RTI International, EdData II</td>
</tr>
<tr>
<td>Philippines</td>
<td>EALAS</td>
<td>Hybrid</td>
<td>Capacity building project; assess learning achievement in curricular domains</td>
<td>Mathematics, Language, Science</td>
<td>Grades 4, 6</td>
<td>2005</td>
<td></td>
<td></td>
<td>UNICEF</td>
</tr>
<tr>
<td>Philippines</td>
<td>Literacy Boost</td>
<td>Hybrid</td>
<td>Program evaluation for specified disadvantaged students and schools; baseline evaluation of early years reading</td>
<td>Literacy</td>
<td>Grades 1, 3</td>
<td>2009</td>
<td>1426 students from 31 schools</td>
<td>Implemented in native languages of the students</td>
<td>Save the Children</td>
</tr>
<tr>
<td>Philippines</td>
<td>Philippine Informal Reading Inventory</td>
<td>National</td>
<td>Literacy</td>
<td>Grades 1 - 6</td>
<td>2004, 2005</td>
<td></td>
<td></td>
<td></td>
<td>Ministry of Education</td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>----------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>PIRLS</td>
<td>International</td>
<td></td>
<td>Literacy</td>
<td>Grade 4</td>
<td>2001, 2006, 2011</td>
<td></td>
<td></td>
<td>IEA</td>
</tr>
<tr>
<td></td>
<td>Core Research Program</td>
<td>National</td>
<td></td>
<td>Literacy, Mathematics, Science, ICT</td>
<td>Preschool-Secondary</td>
<td>2003</td>
<td></td>
<td></td>
<td>Centre for Research in Pedagogy and Practice</td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>--------------------------------</td>
<td>--------------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>--------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Nationwide Assessment</td>
<td>National</td>
<td></td>
<td>Mathematics, Science</td>
<td>Grades 3, 6, 9</td>
<td>2005</td>
<td></td>
<td></td>
<td>Institute for the Promotion of Teaching Science and Technology</td>
</tr>
<tr>
<td>Timor-Leste</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>For policy dialogue</td>
<td>Literacy</td>
<td>Grades 1–3</td>
<td>2009</td>
<td>484 grade 1-3 students in Portuguese, 460 grade 1-3 students in Tetum</td>
<td>Languages of administration: Portuguese, Tetum Tools used as national or system-level diagnostic i.e. representative sample was drawn, and inter-rater reliability was investigated</td>
<td>World Bank</td>
</tr>
<tr>
<td></td>
<td>EALAS</td>
<td>Hybrid</td>
<td>For curriculum evaluation, capacity building project</td>
<td></td>
<td></td>
<td>2005</td>
<td>Due to national conflict, no assessment was implemented. EALAS was used as a tool to develop outcomes and baseline indicators to evaluate the new national curriculum.</td>
<td>UNICEF</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Vietnam</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>To inform an international longitudinal study about childhood poverty (see <a href="http://www.younglives.org.uk">www.younglives.org.uk</a>)</td>
<td>Literacy</td>
<td></td>
<td>2009</td>
<td>2000 students</td>
<td>Language of assessment: Vietnamese Tools used as national or system-level diagnostic i.e. representative sample was drawn, and inter-rater reliability was investigated</td>
<td>DFID, University of Oxford Department for International Development</td>
</tr>
<tr>
<td>Vietnam</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>Baseline to inform multilingual education intervention</td>
<td>Literacy</td>
<td>Grade 3</td>
<td>2010</td>
<td>60 schools</td>
<td>Language of assessment: Vietnamese Tools used for program evaluation i.e. multiple equated forms used to track progress over time</td>
<td>JSDF, Save the Children implemented</td>
</tr>
<tr>
<td>Vietnam</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>Provide information about the literacy of the early elementary students: according assessment competency to international standards for the expects</td>
<td>Literacy</td>
<td>Grade 1 and 3</td>
<td>2013, 2014</td>
<td>1200 students</td>
<td>Language of assessment: Vietnamese tools based on curriculum standards and application of international technical</td>
<td>MOET, WB</td>
</tr>
<tr>
<td>Country</td>
<td>Learning Metric</td>
<td>Type</td>
<td>Purpose</td>
<td>Domains</td>
<td>Year Levels</td>
<td>Year(s)</td>
<td>Sampling</td>
<td>Features</td>
<td>Agencies</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Vietnam (cont.)</td>
<td>EGRA</td>
<td>Hybrid</td>
<td>Provide information about the literacy of the early elementary students; according assessment competency to international standards</td>
<td>Literacy</td>
<td>Grade 1, 2 and 3</td>
<td>2014</td>
<td>1200 students 40 schools</td>
<td>Language of assessment: Vietnamese tools based on curriculum standards and application of international technical</td>
<td>MOET WB</td>
</tr>
<tr>
<td></td>
<td>EALAS</td>
<td>Hybrid</td>
<td>Capacity building project; assess learning achievement in curricular domains</td>
<td>Literacy, Mathematics</td>
<td>Primary</td>
<td>2005</td>
<td></td>
<td></td>
<td>UNICEF</td>
</tr>
<tr>
<td></td>
<td>Literacy Boost</td>
<td>Hybrid</td>
<td>Baseline evaluation of early years reading for ethnic minority students</td>
<td>Literacy, also some Mathematics items</td>
<td>Grade 3</td>
<td>2010</td>
<td>945 students</td>
<td>Languages of assessment: Kinh, Hmong, Thai, Van Kieu, Pako</td>
<td>Save the Children</td>
</tr>
<tr>
<td></td>
<td>PASEC</td>
<td>Regional</td>
<td>Diagnostic evaluation</td>
<td>Literacy, Mathematics</td>
<td>Grades 2, 5</td>
<td>2012-2013</td>
<td>Nationally representative sample: total 5400 students (2700 students/grade 2, 2700 students/grade 5)</td>
<td>Mathematic and Vietnamese Assessment Grade 2 and 5 (input and output)</td>
<td>CONFEMEN Ministry of Education,</td>
</tr>
</tbody>
</table>
SEAMEO EXPERIENCES OF PRIMARY LEARNING METRICS
DESK REVIEW - ACER

SEAMEO countries have less experience with regional assessments. Only with Cambodia, Lao PDR and Vietnam participated in PASEC, due to a similar colonial history with African francophone countries. Globally, the richness of available literature about low and middle income countries’ learning outcomes, policy choices and use of assessment data comes from SACMEQ, PASEC and LLECE (Best et al, 2013). Comparatively, more is known about these issues in Sub-Saharan Africa, Francophone Africa, Latin America and the Caribbean than Southeast Asia, the Pacific, Central and Eastern Europe and the Middle East and North Africa region – which do not have regional assessment programs.

However, results about frequency of assessment types do not consider frequency of implementation (cycles). Most international and regional assessments are administered more frequently and continuously in comparison to what is known about administration of hybrid and national assessments in SEAMEO countries. Consequently, SEAMEO countries that participate in regional and international assessments may have more opportunities to develop technical expertise and institutional capacity for assessment programs and activities. Countries that participate in international assessment programs often develop national assessment programs and other monitoring activities, from increased experience and expertise with the rigorous methodologies of international assessments (Best et al, 2013). Results in the descriptive mapping for hybrid, and national assessments may be incomplete as regards the frequency of administration as current information regarding current activities are often not publicly available, nor detailed. Further research should examine the scope and frequency of these activities with a view to better understand the possibilities for ongoing capacity building and the development of technical expertise with primary learning metrics in SEAMEO countries.

A survey of officials from Ministries of Education and in some instances experts in the field of education and assessment ministries on the use of student assessment for policy and learning improvement (Unesco, 2013b) included questions on the purposes of assessments other than gate-keeping type examinations at the major transitions between levels of formal education (e.g. primary to secondary, secondary to tertiary). The SEAMEO country that responded to the survey were Lao, Myanmar, the Philippines and Thailand. All four countries reported policy design or decision making as a purpose of the assessments. At least two countries also reported policy or programme evaluation, monitoring education qualities and inequalities, or teacher accountability as additional purposes. Results from Best et al (2013) suggest that in the Asia region more generally, large scale assessments were most frequently intended to improve the quality of the education system.

Still, robust policy analysis that examines experiences with learning metrics, and intended and - more importantly - evidence of actual use, is scarce, especially as documentation as regards these activities is not readily available in the public domain. Further case study research should also examine the interactions between the intended aims of participation in these activities to better understand their design, administration and policy choices, the role of different stakeholders as well as evidence of use of results of learning metrics the different stages of the education policy process.
Table 3  Summary of assessment types used in SEAMEO countries

<table>
<thead>
<tr>
<th>Assessment Type</th>
<th>Countries</th>
</tr>
</thead>
</table>
| Hybrid          | Cambodia (EGRA)  
                | Indonesia (Literacy Boost, EALAS)  
                | Lao PDR (EGRA)  
                | Myanmar (EALAS)  
                | Philippines (EGRA, Literacy Boost, EALAS)  
                | Timor Leste (EGRA, EALAS)  
                | Vietnam (EGRA, EALAS, Literacy Boost) |
| International   | Indonesia (TIMSS/PIRLS)  
                | Singapore (TIMSS/PIRLS)  
                | Thailand (TIMSS) |
| Regional        | Cambodia (PASEC)  
                | Lao PDR (PASEC)  
                | Vietnam (PASEC) |
| National        | Cambodia (Learning Assessment System)  
                | Indonesia (Learning Assessment System)  
                | Lao PDR (National Literacy Survey, Assessment of Student Learning Outcomes)  
                | Malaysia (Primary School Achievement Test)  
                | Philippine (Informal Reading Inventory)  
                | Singapore (Core Research Program)  
                | Thailand (Nationwide Assessment)  
                | Vietnam (Reading and Mathematics Assessment Study) |

7. Lessons learned – barriers

This section outlines general barriers to the effective design, implementation and use of primary learning metrics in SEAMEO countries, and more broadly, across low and middle income countries in general. Liberman and Clarke (2012), from their review of World Bank project support for assessment activities in low and middle income countries, identify and describe the following barriers in practice to the effective design, operational activities, and use of learning metrics.

7.1 Design and Implementation

Lack of in-country leadership for operational activities
Liberman and Clarke note that clear and strong leadership is needed at the country level to be able to direct and manage assessment activities, in conjunction with outside technical expertise and other stakeholder support. This finding is also corroborated by results from a systematic review of the impact of large scale assessments on policymaking in low and middle income countries. Weak, and fragmented assessment agencies emerged as a barrier.
to the effective use of assessment data in policymaking (Best et al, 2013) in that continuity of the agency and stability of staff are important for capacity building in assessment activities (ibid).

However, the recommended structure and relative independence of assessment agencies for effective use of assessment data is less clear. Assessment agencies should be linked to the education system/political institutions to be able to respond to relevant policy needs and for feedback loops between the two agencies to develop. At the same time, the independence of assessment agencies from political institutions may help to legitimise assessment activities in the long-term with the public. Lastly, stable assessment agencies can help to insulate assessment activities from political instability and regime change, as in the case of a country that had four different education ministers during the implementation of a project (Liberman & Clarke, 2012).

Misalignment of project indicators and project inputs
Project inputs can be broadly defined as instrument design, sampling, project resources and technical expertise, among other inputs. These inputs should align with project indicators, or goals, purposes and expected outcomes of learning metrics. As discussed, less is currently known about purposes and goals of regional and national primary learning metrics in SEAMEO countries. These should be clarified in order to ensure that all project inputs align with project indicators and assessment activities.

Inadequate funding of assessment activities
Insufficient funds were also cited as a barrier to the effective use of assessment data (Best et al, 2013), not only to undertake appropriate operational activities, but also to undertake further in-depth analyses that may help to address policy concerns. Countries that addressed this barrier increased funding for the assessment agency, legislated funding for the assessment body, and aligned assessment activities to donor goals and priorities to ensure funding from donors.

Pressure from assessments
Unesco (2013b) reported perceived pressures of assessments through teacher unions or parents and the wider community as well as on students having to sit these assessments, although these pressures were not differentiated according to the types and purposes of the assessments. However, it is clear that with the move to greater accountability and transparency, the pressures on students and teachers and the time taken to undertake various assessments for different purposes has increased.

7.2 Policy Use

The following are barriers to the effective use of learning metrics, or assessments for policy and decision making (Best et al 2013).

Technical soundness
The soundness of the assessment program or learning metrics is the principal barrier to effective use for policymaking, as low technical soundness of a learning metric can lead to inappropriate interpretations and use of results. Aspects related to the technical quality include:

- Technical capacity of staff
- Assessments not well designed for meaningful comparisons and assessment goals
- Assessments are not relevant to policy of contextual concerns; nor do they provide information regarding the extent, or amount of inputs for policy choices
- Assessments are infrequently implemented and results are not comparable over time
- Low quality sampling and administration

Dissemination activities
Dissemination activities can serve as a barrier to the use of assessments and learning metrics for policymaking and decision making specifically when dissemination activities do not consider the public and larger civil society, and/or are inadequate or inappropriate. (Best et al, 2013). Dissemination to the public should be appropriate for stakeholder needs, and should also address the public’s expectations regarding the information gained from assessment programs. For example, inappropriate expectations of data and learning metrics by the public can lead policy makers to inappropriately use results to meet public expectations (ibid).

Systems Issues
The systematic review of assessment impact on policymaking noted a lack of information about system level factors’ relationship to the use of data (ibid). These systems include the wider political, social and economic systems as well as the overall functioning of the education system. Liberman and Clarke (2012) identify system issues that served as barriers to assessment activities and learning metrics. National crises or political instability often served as barriers to effective assessment activities. Additionally, teachers unions and parental organisations often oppose the introduction of new learning metrics.

8. Innovative approaches - facilitators

Overall, there is a lack of evidence that demonstrates the use of data from learning metrics and assessment programs for decision making and policymaking in low and middle income countries (Best et al, 2013). Considering the evidence that does link learning metrics to policy use, information from these activities are used less in policy formulation, than used to identify issues, implement or monitor policies. Consequently, less is known about the factors or levers that successfully drive the effective use of data and information from learning metrics, regardless of type of assessment program, or level of education.

However, from the available evidence, policymakers in low and middle income countries most often used assessment data for the following reforms and policies:

Overall, the most frequent education policies resulting from the use of assessment data were system-level policies regarding curriculum standards and reform, performance standards, and assessment policies (ibid). Resource allocation policies most often aimed to increase teacher quality and teaching materials through in-service professional development, improved teacher training, and textbook reform. Additionally, resource allocation policies also reformed school funding. Use of learning metrics and assessment data resulted less frequently in teaching and learning policies, although, policies for student-oriented pedagogy and in-class learning strategies were documented (ibid).
The following approaches have served as facilitators to the effective design, administration and use of assessments.

8.1 Dissemination

As discussed, dissemination can serve as a barrier or facilitator for effective assessment activities. Effective dissemination to the media and to the public has served as an accountability lever for policy enactment. However, there may be many political sensitivities about dissemination to the media and broader public, as well as the possibility for unrealistic expectations of assessments and policy reforms, by the public.

Appropriate dissemination to stakeholders includes broad dissemination to stakeholders at various levels, as well as targeted dissemination appropriate to the stakeholder group (ibid). For example, Thailand widely disseminated results from TIMSS and PISA by making reports available online, providing key national and international stakeholders with reports, press releases, summaries and presentations to stakeholders, organisation of conferences to discuss results, and well as providing feedback to schools (UNESCO, 2013b). While feedback to schools does not necessarily make a difference to the impact of an assessment program, it assists with its longevity as stakeholder groups, including parents, teachers, teacher unions are more convinced of the usefulness of a program from which they gain some information. However, results from TIMMS and PISA may have been lower-stakes as compared to other assessments, due to the sampling of the assessment.

Examples of differentiated assessment may include: the creation of parent and student guides to report results, the inclusion of test items and scoring rubrics in teacher reports to help teachers create their own classroom based assessments, or face to face meetings with policymakers (Best et al, 2013). Dissemination should also be timely, as appropriate for the assessment (Liberman & Clarke, 2012).

8.2 Soundness of program

The soundness of the assessment program can serve either as a barrier or a facilitator. Important factors for the soundness of the program include: credible and reliable data; defined assessment frameworks and standards to help with interpretation of results that are realistic for the context; regular administration of the assessment.

Additionally, key stakeholders should be involved in the initial development of the assessment, to help support buy-in and ownership of the assessment and assessment data (Liberman & Clarke, 2012). For example, PASEC, as well as SACMEQ, have prioritised regional policy concerns in the design of the assessment which made it more likely that results would be used in policy-making (Best et al, 2013). Regional policy concerns have included grade repetition, double-shift classrooms and teacher contracts, for example.

In order to better address the current barriers to the effective development and use of learning metrics, and use knowledge of effective facilitators, some possible frameworks could help guide stakeholders to better align assessment activities within education systems. For example, Clarke (2011) has developed an initial policy framework for researchers, policymakers, donors and other stakeholders to use to evaluate assessment systems and identify areas for improvement. The framework proposes several indicators along key policy domains that can be used to improve institutional capacity for assessments, and to use
assessments for improving educational outcomes: enabling context, system alignment and the quality of assessments.

9. Summary

In this section, the findings of this desk review are summarised in terms of objectives, target populations, content, logistics and frequency of assessment and facilitators.

9.1 Objectives

The following objectives and purposes for primary learning metrics in SEAMEO were identified in this desk review:

- Assess learning achievement, overall and subgroups
- Assessment of individual student performance
- Baseline information
- Benchmarking to international standards
- Capacity development in educational assessment design, analysis and reporting
- Comparisons over time
- Curriculum evaluation
- Diagnostic information
- Evaluation of program/initiatives
- Inform teacher training
- Inform instruction
- Inform policy making/policy dialogue
- Inform teacher professional development
- Monitor educational quality
- System-level monitoring of educational equity and quality

While most of the purposes above can be achieved through national assessments, benchmarking to international standards, for example, can only be achieved through cross-national assessments. However, the usefulness of benchmarks for low or medium income countries through current large-scale assessments has been questioned (e.g. Kellaghan and Greaney, 2001). More specifically, the overall low level of performance of low and middle income countries in those international tests combined with a very low variance means that data analyses for these countries rarely provide insights into a) what students actually are able to do in the assessed content areas and b) factors that are associated with differences between higher and lower performing students.

International agencies are aware of these restrictions as illustrated by their attempt at catering to a greater extent to the needs of economically developing countries, for example, through "Pre-PIRLS" by the IEA or "PISA for Development" by the OECD.

Still, the strength of regional assessments such as LLECE or SACMEQ stems only partly from a specific tailoring of the performance tests. An even greater importance for the success of these regional assessments (see Best et al., 2013) stems from the greater homogeneity of participating countries in terms of culture and geography and political set-up and educational challenges. This homogeneity enables a regional assessment to address
more specifically the needs and questions that are relevant for the continuous improvement of the education systems in the participating countries. Having greater homogeneity along these lines, then, makes for arguably more informative comparisons as less of the differences in performance between countries can be assigned to contextual or cultural differences. This, in turn, results in greater information value in terms of examining the possible impact of differences, for example, in teacher training, instructional resources and practices on differences in learning outcomes.

9.2 Target populations

Target populations of the reviewed primary learning metrics cover the whole range of primary education from Grade 1 to Grade 6. The lower Grades, however, are mainly covered through EGRA, which allows modification to the assessment that restricts the ability to make comparisons across countries. Wagner et al (2013) have called for the assessment of learning outcomes to occur at the lower level of primary education as children in economically developing countries often have already left formal schooling by the end of primary school. The authors also argue that information is needed earlier to increase the effectiveness of any interventions or targeted instructional strategies or sources. However, the benefits of an assessment programme at the earlier stages of primary schooling has to be balanced against the greater restrictions of the time and complexity of the assessment and its implications on the accuracy of measurement, not only for the cognitive but also the attitudinal aspects of the assessment.

9.3 Content

Most of the reviewed primary learning metrics cover literacy and mathematics as the content domain and, to a lesser extent, science. This reflects the importance of the skills that are considered to be the basis for the learning of further educational content and subsequent participation in the workforce.

At a recent Regional Expert Reference Group meeting, strong interest was expressed in two additional content areas, namely writing and global citizenship. Cross-national assessment in these content areas are possible as evidenced by the Written Composition Study (http://www.iea.nl/written_composition_study.html) and the International Civics and Citizenship Study (ICCS; http://www.iea.nl/iccs_2009.html), both undertaken by the International Association for the Evaluation of Educational Achievement (IEA).

The Written Composition Study actually was undertaken at the primary school level as it included 10-12 year olds as the youngest of the three target populations. One SEAMEO country, namely Thailand, was one of the sixteen countries that participated in the study which was undertaken in 1984-85. One of the major challenges of the survey was to develop coding schemes and to train coders in their use to mark students' writing in a reliable and comparable fashion across languages.

In 2009, the International Civics and Citizenship Study had eighth grade as its target population, provided that the mean age at the time of testing was at least 13.5 years. This was considered the youngest age at which students in participating countries could be realistically expected to have had sufficient exposure to the topics covered in the assessment. Indonesia and Thailand were the two SEAMEO countries that participated in
ICCS. Together with Korea, Chinese Taipei and Hong Kong SAR, they formed an Asian module that assessed content, attitudes and values specific to these countries. ICCS will next be conducted in 2016 and countries are currently encouraged to participate.

9.4 Logistics and frequency of assessment

This desk review has shown that the primary learning metrics undertaken in individual SEAMEO countries are mainly undertaken by the Ministry of Education and vary in terms of the interval between assessments.

The regional assessments in South East Africa and South America (i.e. SACMEQ and LLECE) are undertaken in five- to six-year intervals. Assessment activities are coordinated through a central secretariat.

9.5 Facilitators of successful large-scale assessments

Based mainly on the results of a systematic report on this topic (Best et al. 2013), the list below summarizes the barriers and facilitators of the impact of large scale assessments on education policy discussed in greater detail in Sections 7 and 8 above.

Technical soundness
Large scale assessments that result in credible and reliable data which are relevant to the educational policy goals of the participating countries are more effective in terms of their impact than others. Assessments that are undertaken at regular intervals have greater information value and are more relevant. Technical soundness of an assessment programme includes high quality sampling and standardised administration which is checked by quality assurance monitors during field-work.

Dissemination activities
The success of a large-scale assessment is ongoing and regular dissemination activities to different stakeholders. More successful assessments communicate in different modes and with different complexities to the general public, schools, parents, teachers, unions as well as education policy-makers. Such ongoing dissemination and communication activities facilitate the ongoing buy-in and continued interest of different lobby groups in the results of the assessment programme and assist with raising the importance of the assessment to ensure continued support by the different stakeholders.

Funding
More successful assessment programmes have worked to ensure funding for operational activities and further analyses, either through legislating assessment costs in the budget and/or through an alignment with donor goals and priorities.

Strong leadership
Buy-in and support at the highest level of leadership both within countries and across countries are essential for a successful assessment program. One essential component of the success of the SACMEQ regional assessment has been the involvement of education
policy-makers in the design of the assessment materials to ensure that information produced by the assessment is relevant to the educational policy making process.

Strong leadership includes decisions and subsequent commitment to the agreed purposes and central questions to be answered by the assessment. It also requires decisions regarding setting priorities in terms of the stages of the education policy cycle, such as agenda setting, policy formulation, policy implementation and monitoring and policy evaluation that the assessment is meant to inform. In a regional or international assessment, strong leadership includes direction setting and agreement on these issues at the highest level of government of the countries involved.

It is probably the drive and momentum gained from a strong and visionary leadership for all the aforementioned facilitators that - combined with hard work at all levels of the implementation of the assessment - that will make a possible regional assessment programme in SEAMEO countries a success.
REFERENCES


FTI (2009). ‘Eligible Countries Engaged in Externally Provided Learning Assessments’ FTI_countries_engaged_in_learning_assessments.xlsx


