



Evaluation of Southeast Asia- Primary Learning Metrics

Inception Report

8 October 2020

Cambridge Education
22 Station Road
Cambridge CB1 2JD
United Kingdom

T +44 (0)1223 463500
camb-ed.com

Evaluation of Southeast Asia- Primary Learning Metrics

Inception Report

7th October 2020

Issue and Revision Record

Revision	Date	Originator	Checker	Approver	Description
1	15.05.20	Dr Rachel Outhred	Alice Wright	Emma Mba	Inception Report (Draft 1)
2	02.07.20	Dr Rachel Outhred	Alice Wright	Emma Mba	Inception Report
3	07.10.20	Dr Rachel Outhred	Alice Wright	Emma Mba	Inception Report (revised)

Document reference:

Information class: Standard

This document is issued for the party which commissioned it and for specific purposes connected with the above-captioned project only. It should not be relied upon by any other party or used for any other purpose.

We accept no responsibility for the consequences of this document being relied upon by any other party, or being used for any other purpose, or containing any error or omission which is due to an error or omission in data supplied to us by other parties.

This document contains confidential information and proprietary intellectual property. It should not be shown to other parties without consent from us and from the party which commissioned it.

Contents

List of abbreviations	v
1 Introduction	7
1.1 Overview	7
1.2 Inception process	7
1.3 Inception report structure	8
2 Context of the Evaluation of the Southeast Asia Primary Learning Metrics (SEA-PLM) Programme in Southeast Asia	10
2.1 Regional and country context	10
2.1.1 Education in Southeast Asia	10
2.1.2 Access and equity across the region	10
2.1.3 Learning and equity across the region	13
2.1.4 Education System Quality Indicators in SEA-PLM participating countries	13
2.1.5 Learning assessment programmes and Southeast Asia	13
2.2 Object of the evaluation - SEA-PLM	16
2.2.1 Emergence of SEA-PLM	16
2.2.2 Governance structure of SEA-PLM	17
2.2.3 Technical structure of SEA-PLM	18
2.3 Other partner support to improving access to quality education for all in the region	19
2.4 SEA PLM Logframe	19
2.5 Reconstructed and refined Theory of Change (ToC)	19
2.5.1 Assumptions	22
3 Evaluation purpose	24
3.1 Type of evaluation (why being done at this time)	24
3.2 Primary and secondary audiences	24
3.3 Duty bearers and rights holders (and their involvement)	24
3.4 Evaluation objectives	25
3.5 Scope of the evaluation	25
4 The evaluation framework	26
4.1 Summary of the literature review	26
4.2 Any adaptations from the initial ToR	29
4.3 Evaluation criteria and questions	29
4.3.1 Prioritization of the evaluation questions	30
4.4 Hypothesis to be tested	31

5	Methodology	32
5.1	Conceptual framework	32
5.2	Evaluation approach	33
5.3	Data	35
5.3.1	Data collection methods	35
5.3.2	Data collection instruments	37
5.3.3	Data collection consistency and quality	38
5.4	Data analysis	38
5.4.1	Qualitative Comparative Analysis	38
5.4.2	Synthesis	38
5.4.3	Triangulation, bias and informed judgements	38
5.5	Equity and gender	39
5.6	Ethics	39
5.7	Limitations	40
6	Workplan	41
6.1	Evaluation management and quality assurance	44
6.2	Roles and responsibilities	44
6.3	Communication and coordination	45
6.4	Logistics and support needed from UNICEF	45
6.5	Deliverables and reporting	46
6.6	Evaluation dissemination and utilization plan	47
7	Risk management and risk mitigation	48
7.1	Realities and scenarios - COVID-19	49
7.1.1	Mitigating the challenges of remote research	51
7.2	Additional risks introduced by COVID-19	52
8	Bibliography	54
	Annexes	57
	Annex A - Terms of Reference	57
	Annex B - Learning and equity across Southeast Asia	70
	Annex C - Education System Quality Indicators in SEA-PLM participating countries graphs	72
	Annex D - Partner support to education in SEA-PLM participating countries	76
	Annex E - Evaluation matrix	78
	Annex F - Logframe	85
	Annex G - Refining the ToC	88
	Annex H - Data collection tools	97
	Annex I - Interview guide	102
	Annex J - Literature review	104

1	Introduction	1
1.1	The purpose and origin of this literature review	1
1.2	The evaluation of SEA-PLM for which this review is undertaken	1
1.3	Contribution of the literature review towards the evaluation of SEA-PLM	1
2	Methodology	3
3	Findings	4
3.1	Introduction to international and regional learning assessments	4
3.2	Evidence of the impact of the international and regional learning assessments	6
3.2.1	PISA	6
3.2.2	PISA Shock	6
3.2.3	Wider PISA Impact	7
3.2.4	Criticisms of PISA	8
3.2.5	SACMEQ	9
3.2.6	PASEC	13
3.3	Enablers and barriers to achieving impact	15
3.3.1	The government and the external, institutional environment in the member country	15
3.3.2	The skills and characteristics of the institutions and individuals who lead the assessments	16
3.3.3	Factors integral to the assessment, including the rules and commitments associated with participating	16
3.3.4	The nature and analysis of the learning assessment findings	17
3.3.5	The approach to disseminating the findings	18
3.3.6	Responding to and acting on the findings	19
4	Summary of findings	20
5	Bibliography	23

Tables

Table 1	Trends across access, enrolment, schooling life expectancy, repetition and graduation indicators (2013 to 2019), using spotlight rating	12
Table 2	Participation in cross-country assessments in the Southeast Asia region, 2005-2019	14
Table 3	Example of how the evaluation matrix will be used, with reference to evaluation question one	34
Table 4	Risks, probability, impact and mitigation	48
Table 5	Implications of COVID: Scenarios A and B	50
Table 6	SEA-PLM Activities	93
Table 7	Interview items by stakeholder group	97

Figures

Figure 1	International and regional assessment participation in the Asia Pacific between 2005 and 2015	14
Figure 2	SEA-PLM Reconstructed ToC	21
Figure 3	Key functions of process evaluation, from Moore et al, 2014	32
Figure 4	The three components underpinning the evaluation methodology	33
Figure 5	Workplan	42
Figure 6	Border closures in Asia	49
Figure 7	Trends in Grade 8 Maths Proficiency, TIMMS	70
Figure 8	Trends in Grade 8 Science Proficiency, TIMMS	70
Figure 9	Average Proficiency in Maths, by Gender (TIMMS 2015)	71
Figure 10	Average Proficiency in Science, by Gender (TIMMS 2015)	71
Figure 11	Primary school age population (in millions) 2019	73
Figure 12	Pupil-teacher ratio in primary education	73
Figure 13	Percentage of trained primary teachers	73
Figure 14	Proportion of primary schools with basic handwashing facilities	73
Figure 15	Primary Gross enrolment ratio, Primary, Gender parity Index	74
Figure 16	Primary school life expectancy, by gender 2017	74
Figure 17	Survival rate to the last grade of primary education, by gender	75
Figure 18	Number of repeaters across primary grades, by gender	75
Figure 19	Percent of enrolled primary school students who are overage and underage	75

List of abbreviations

ACER	Australian Council for Educational Research
ASEAN	Association of Southeast Asian Nations
CLMV	Cambodia, Lao PDR, Myanmar and Vietnam
DAC	Development Evaluation Criteria
DIFF	Differential Item Functioning
EAPRO	East Asia and Pacific Regional Office

ECD	Early Childhood Development
ETS	Education Testing Service
GAML	Global Alliance for Monitoring Learning
GEROS	Global Evaluation Reports Oversight System
GPE	Global Partnership for Education
GOM	Government of Myanmar
IEA	International Association for the Evaluation of Educational Achievement
LMIC	Low- and Middle-Income countries
MOE	Ministry of Education
ODA	Overseas Development Assistance
OOSC	Out of School Children
SDG	Sustainable Development Goal
SEAMEO	Southeast Asian Ministers of Education Organization
SEA-PLM	Southeast Asia Primary Learning Metrics
ToC	Theory of Change
UNEG	United Nations Evaluation Group
UIS	UNESCO Institute of Statistics
VfM	Value for Money

1 Introduction

1.1 Overview

Around the world, large-scale survey assessments of learning are conducted at national, regional and international levels. These surveys are to produce evidence of learning achievement taking place on a basis which enables statistically valid comparisons to be made over time and across different contexts. Such surveys are instruments intended for the measurement of progress, to ascertain the effectiveness of education systems, to monitor progress towards equity, and to guide investment and effort.

In Southeast Asia, the Southeast Asia Primary Learning Metrics (SEA-PLM) programme has been developed and piloted in six countries. SEA-PLM is a sample-based survey of learning at Grade 5 level. It has been developed with the technical assistance of the Australian Council for Educational Research (ACER).

UNICEF has been a key participant and funder of SEA-PLM. The first reporting cycle of data will be in 2020. Prior to investing in the longer-term expansion and institutionalisation of the survey, UNICEF at the present juncture wishes to take stock of the experience to date. To do so, UNICEF wishes to commission an evaluation of SEA-PLM, to guide itself, and other stakeholders, for the way forward.

The Southeast Asian Ministers of Education Organization (SEAMEO) is also a key stakeholder in the SEA-PLM. The purpose of SEAMEO is to promote cooperation in education, science and culture in the Southeast Asian region. UNICEF provides support to the SEAMEO Secretariat. The Association of Southeast Asian Nations (ASEAN) and the Global Alliance for Monitoring Learning (GAML) are also identified in the Terms of Reference as beneficiaries of the strengthening of regional alignment and integration, with GAML responsible for developing measures of learning towards Sustainable Development Goal 4. With its regional basis, and the opportunities it embodies for inter-country comparisons and standards, SEA-PLM is in line with the philosophy of strengthening such regional alliances. Participating countries are also key partners in SEA-PLM. Participating countries include Cambodia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines and Viet Nam.

In its simplest form, the Theory of Change (ToC) for national, regional and international assessments is that if data on the learning levels of students is available, better decisions regarding how to resource classrooms, train teachers and enact impactful policies will be made. The technical complexities of measuring learning and analysing the factors associated with learning success require a set of specific technical expertise, not always present in Low- and Middle-Income countries (LMIC). Furthermore, the capacity to assess, access and use learning outcome data to inform decisions is often lacking and in many countries and institutions opening up political space to support evidence-based decision making is a complex and long-term challenge.

In this way the SEA-PLM addresses both technical and political complexities, across a variety of diverse countries in partnership with a diverse set of actors. Understanding how the programme is working thus far to address technical and political realities towards individual behaviour change, institutional reform and ultimately learning outcomes, is crucial at this stage of SEA-PLM implementation.

In order for SEA-PLM to be effective and sustainable, reflection and possible adaptation of the programme delivery model or structure should be undertaken. This evaluation will provide robust data to inform decisions regarding SEA-PLM.

1.2 Inception process

The inception phase for this evaluation ends with approval of a final inception report and the delivery of a desk study of SEA-PLM documentation and a literature review. The desk study and literature review provide background understanding of the SEA-PLM and context for the evaluation, in relation to good practice and experience in regional and global large-scale learning assessments. Both the

study and the review have also fed into the reconstructed Theory of Change (ToC) developed in this inception report, to be used as the guiding framework for the evaluation.

Literature search and review

A focused literature search and review has been undertaken by the evaluation team (See Annex F Literature Review). The literature review provides background understanding of the SEA-PLM and provides context for the evaluation, in relation to good practice and experience in regional and global large-scale learning assessments.

The literature review may be revisited and finalised following feedback.

Desk review of SEA-PLM documents

The purpose of the desk review is to understand the programme's history, inform the methodology and provide context and data for the evaluation. While the literature search and review is global, and the documentation for it located and selected by the evaluation team, the desk review of SEA-PLM was restricted to key documents on SEA-PLM, provided by UNICEF EAPRO. The documentation relates to the eleven countries of Southeast Asia.

The desk review also serves as useful background for the key stakeholder interviews and country case studies and informs the inception workshop and the ToC, as mentioned above. The desk study is incorporated integrally within the inception report, rather than as a separate element.

Process for the approval of the inception report

Following the delivery of a draft inception report, an inception workshop was undertaken as an effective opportunity for consultation, review, refinement and agreement of the ToC, the workplan and reporting requirements, as well as practical aspects of management and working practicalities of the evaluation assignment.

The inception workshop was virtual and included a presentation of the evaluation methodology, followed by the presentation of the ToC. Substantive inputs and feedback from the evaluation Reference Group were elicited, and the ToC was adapted and refined on this basis. These refinements are outlined in Section 2.5.

In addition, the evaluation team refined and revised this inception report through the consultations carried out during the workshop and based on the written feedback by the Reference Group. This final inception report addresses stakeholder comments and incorporates feedback, where appropriate.

The evaluation inception workshop used Mentimeter software to gather data on stakeholder perceptions on the importance, timing and scope of the evaluation. Participants were asked to state the extent to which they agree with specific statements on a scale of 1-5. On average participants strongly agreed that the evaluation is very important (average 4.5), that it is a good time within the SEA-PLM programme to conduct an evaluation (3.9), that it is very important to assess the governance (4.5) technical (4.5), managerial (4.3) and funding (4.2) aspects of SEA-PLM.

1.3 Inception report structure

This inception report outlines the context of the evaluation of SEA-PLM, the evaluation purpose, the evaluation overview, the methodology, the workplan and risk management.

Section 2, the context of the evaluation of the SEA-PLM (page 9) places SEA-PLM within the regional context and summarises the contexts of countries within the region, including access, quality and equity and trends in education indicators. An overview of the SEA-PLM programme is outlined, along with partner support to improving access to quality education for all in the region. The SEA-PLM logframe is presented, along with a reconstructed ToC.

Section 3 (page 23) outlines the purpose of the evaluation, including a description of the type of evaluation, identification of the primary and secondary audiences, how users, duty bearers and rights holders have been involved in the evaluation and the objectives and scope of the evaluation.

Section 4 (page 25) provides an overview of the evaluation, summarising the literature review, any adaptations from the initial ToR, outlining the evaluation criteria and questions and putting forward the hypothesis to be tested.

The methodology for conducting the evaluation is laid out in Section 5 (page 31). This includes the conceptual framework, the evaluation approach, design and methodology, the sampling approach, data sources, methods and instruments, data analysis, equity and gender considerations, ethics and methodological limitations.

The evaluation workplan is described in Section 6 (page 41). This includes the management arrangements for the evaluation, quality assurances protocols, roles and responsibilities, communication and coordination, logistics, deliverables and reporting and the evaluation dissemination and utilization plan.

The final section of the report (Section 7 page 48), outlines the risk management and risk mitigation approach for this evaluation. Importantly, the section addresses the impact of the current global pandemic on the implementation of the evaluation and provides several options for quality implementation based on a range of potential scenarios, in these uncertain times.

2 Context of the Evaluation of the Southeast Asia Primary Learning Metrics (SEA-PLM) Programme in Southeast Asia

SEA-PLM aims to support Southeast Asia Ministries of Education Organization (SEAMEO) member countries to measure the status of learning achievement in reading, writing, mathematics and global citizenship. (ACER, SEAMEO & UNICEF, 2017).

The overall goal of the SEA-PLM Programme is that by 2025, all Southeast Asia countries have regional standardized quality measures to assess learning outcomes for grade 5 students (roughly 10-11 years of age) in the domains of reading, writing, math and global citizenship and consistently use these to inform improvement within the education sector (SEA-PLM SEAMEO UNICEF, 2019b, p. 16).

SEA-PLM is in the last stages of the first round of the first regional large-scale assessment in the Southeast Asian Region. The results for SEA-PLM 2019 will be reported and disseminated in late 2020, supporting the use of the data through to 2021 (UNICEF, 2019). For this reason, it is necessary for partners to explore the strength and weaknesses of the first assessment round before engaging in the second round of SEA-PLM. This includes a shift from a biannual plan to a five year commitment requiring increased stability, visibility and ownership of the project (UNICEF, 2019, p. 4).

This section of the report outlines the regional and country contexts, including key education indicators across the areas of access, quality and equity. The objective of the evaluation, partner support to education in the region and the SEA-PLM logframe and reconstructed ToC are presented.

2.1 Regional and country context

The East Asia and Pacific region houses one-third of the global population and more than a quarter of the world's children. The geography, culture, economic and political systems diverse significantly across the region. Disparities in wealth between nations can be observed and in many countries disparities in wealth within countries also diverges. This also applies to ethnic and linguistic diversity in the region (UNICEF EAPRO, 2020).

2.1.1 Education in Southeast Asia

Increased enrolment, retention and completion rates and decreased gender gaps have been observed within some countries in the region, however many children remain out of school (OOS) or lack access to quality education (UNICEF EAPRO, 2020).

2.1.2 Access and equity across the region

Across the eleven countries in Southeast Asia, recent progress in improving access, quality and equity varies. As can be seen in Table 1 Trends across access, enrolment, schooling life expectancy, repetition and graduation indicators (2013 to 2019), using spotlight rating, (based on available UIS data) Brunei Darussalam's progress is constrained across many indicators, Cambodia is making positive progress in reducing overage enrolment and gender equity, and Indonesia is making positive process in reducing the percentage of Out-of-School Children (OOSC). In Indonesia primary school life expectancy has improved for boys, while secondary school life expectancy has improved for girls. Laos is making positive progress in reducing overage and underage enrolment, increasing access to Early Childhood Development (ECD), gender parity in primary enrolment and secondary school life expectancy. Malaysia has reduced overage enrolment and primary school life expectancy for boys, with improvements for girls remaining stagnant or small. Myanmar has made advancements in reducing the rate of OOSC, increasing school life expectancy at the primary and secondary levels and improving the graduation ratio at the primary level. The Philippines has reduced overage enrolment

and improved secondary school life expectancy. Thailand has made improvements in access, school life expectancy at the primary and secondary levels, reduced repetition in primary school and graduation ratio at the primary level. Timor-Leste has reduced the rate of OOSC, improved gender parity in enrolments at the primary level, improved school life expectancy at the secondary level, reduced the rate of primary repetition and improved primary graduation ratios. Based on the available data, Vietnam has improved across the full range of indicators, with the expectations of reducing underage enrolment, school life expectancy for primary boys and improving gross graduation ratios (where indicators have remained reasonably stable).

Gross graduation ratios remain reasonably stable in Brunei Darussalam, Indonesia, Malaysia, the Philippines and Vietnam because starting ratios were near or at 100 percent in 2013.

There is very little comparable data available for these indicators in Singapore.

Table 1 Trends across access, enrolment, schooling life expectancy, repetition and graduation indicators (2013 to 2019), using stoplight rating¹

	Rate of OOSC of primary school age, both sexes (%)	Rate of OOSC of primary school age, girls (%)	Rate of OOSC of primary school age, boys (%)	Percentage of students enrolled in primary education who are over-age,	Percentage of students enrolled in primary education who are under-age,	Gross enrolment ratio, early childhood educational programmes, both	Gross enrolment ratio, early childhood educational development	Gross enrolment ratio, early childhood educational development	Gross primary enrolment ratio, both sexes (%)	Gross enrolment ratio, primary, gender parity index	School life expectancy, primary, female (years)	School life expectancy, primary, male (years)	School life expectancy, secondary, female (years)	School life expectancy, secondary, male (years)	Percentage of repeaters in primary education, all grades (%)	Gross graduation ratio from primary education (%)
Brunei Darussalam	Red	Grey	Red	Grey	Red	Grey	Grey	Grey	Red	Green	Red	Red	Red	Red	Red	Yellow
Cambodia	Red	Red	Red	Green	Yellow	Grey	Grey	Grey	Red	Green	Red	Red	Grey	Grey	Red	Red
Indonesia	Green	Red	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Red	Green	Green	Yellow	Green	Yellow
Laos	Red	Red	Red	Green	Green	Green	Green	Green	Red	Green	Red	Red	Green	Green	Green	Yellow
Malaysia	Yellow	Grey	Red	Green	Grey	Grey	Grey	Grey	Yellow	Yellow	Green	Green	Yellow	Yellow	Grey	Red
Myanmar	Green	Grey	Red	Red	Grey	Grey	Grey	Grey	Green	Red	Green	Green	Green	Green	Green	Green
Philippines	Red	Red	Yellow	Green	Red	Grey	Grey	Grey	Red	Yellow	Red	Red	Green	Green	Yellow	Yellow
Singapore	Grey	Grey	Grey	Grey	Yellow	Grey	Grey	Grey	Yellow	Grey	Grey	Grey	Grey	Grey	Grey	Grey
Thailand	Grey	Grey	Grey	Green	Red	Red	Red	Red	Green	Green	Green	Green	Green	Green	Green	Green
Timor-Leste	Green	Green	Green	Red	Grey	Grey	Grey	Grey	Red	Grey	Red	Red	Green	Green	Green	Grey
Vietnam	Grey	Grey	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Grey	Grey	Green	Yellow

¹ Green stoplight rating denotes improvements, yellow denotes stagnation or small movements in trends and red denotes deterioration. Grey reflects an absence of data. Source UIS data 2013-2020. Access refers to access to schooling. Enrolment refers to enrolment in a school. Schooling life expectancy refers to the number of years of schooling an average children receives. Repetition refers to repeating grades. Graduation refers to graduating from a sub-sector of schooling.

2.1.3 Learning and equity across the region

Underscoring the impetus behind SEA-PLM, data on learning outcomes are not available for many countries in the region, particularly in reading and writing. Of the eleven countries, only Singapore has participated in PIRLS and Malaysia, Singapore and Thailand have participated in the Grade 8 TIMMS (Mullis, Martin, Foy, & Hooper, 2016).

Annex B outlines the trends in Mathematics and Science across participating countries in the region. In summary, average proficiency in learning has consistently been within the Advanced International benchmark for Singapore, average proficiency in learning has consistently been within the Low International Benchmark for Thailand and for Malaysia, proficiency has moved from the Intermediate International benchmark to the Low Benchmark over the years.

Girls perform higher than boys in each of the three participating countries in Mathematics and perform higher in Science in Malaysia and Thailand, while boys perform higher in Science in Singapore.

2.1.4 Education System Quality Indicators in SEA-PLM participating countries

Of these eleven countries, six countries have participated in SEA-PLM, including Cambodia, the People's Democratic Republic of Lao, Malaysia, Myanmar, Philippines and Vietnam.

These six education systems are diverse in terms of size, quality and equity. The primary school age population in Lao PDR and Viet Nam, for example, are over 7 million, while for Cambodia the primary school age population is 2 million. The pupil-teacher ratio in primary education ranges from approximately 12 in Malaysia to approximately 43 in Cambodia. Similarly, approximately 93 percent of schools in Malaysia have basic handwashing facilities, while in Cambodia just under half of all schools have these facilities. For other education quality indicators, the countries diverge less. For example, the percentage of trained teachers across the countries are reasonably similar, as is the Gender Parity Index for primary school enrolment.

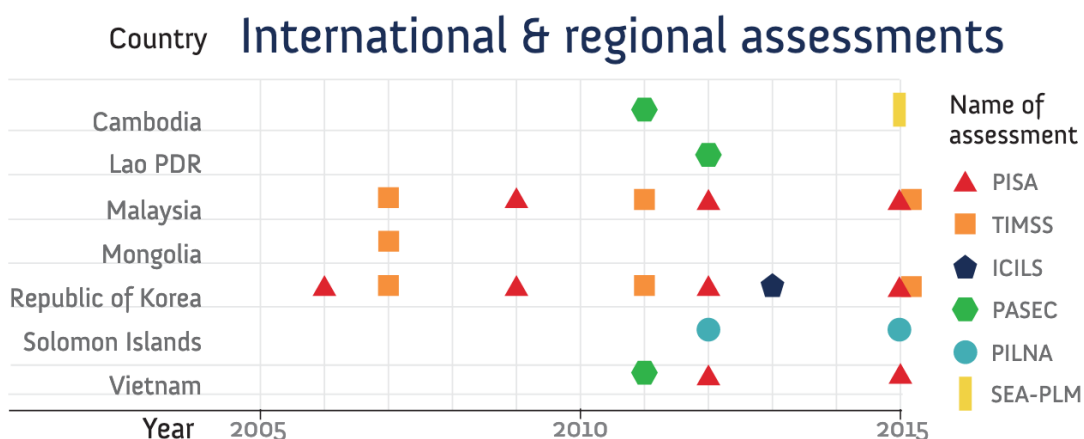
2.1.5 Learning assessment programmes and Southeast Asia

The need for a diverse set of assessments has been recognised globally since the shift from a focus on access to learning in the global development goals. Regional assessment initiatives include the Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ), the Latin American Laboratory for Assessment of the Quality of Education (LLECE), and the CONFEMEN Programme for the Analysis of Education Systems (PASEC). The Russian Education Aid for Development (READ) has also been utilized across a number of countries in Eastern Europe. The World Bank initiatives for partner countries such as the Systems Approach for Better Education Results (SABER) provides assessment and comparative data on education policies and institutions in the Central Asian region and in parts of Africa (Sankar, 2015).

As stated above, of the eleven countries in the Southeast Asian region, Singapore, Malaysia and Thailand have participated in global assessment programmes. Singapore has participated in PIRLS and Malaysia, Singapore and Thailand have participated in the Grade 8 TIMMS (Mullis, Martin, Foy, & Hooper, 2016). Cambodia also participated in PASEC in 2011 and Lao in 2012. Malaysia participated in PISA in 2009 and in 2012, along with Vietnam.

The Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP) Secretariat at UNESCO Bangkok surveyed 13 countries in the region in order to learn more about regional participation in national, regional and international assessments between 2005 and 2015. (NEQMAP Secretariat, 2015).

Figure 1 International and regional assessment participation in the Asia Pacific between 2005 and 2015



SEA-PLM is the first regional large-scale learning assessment to be specifically designed for the Southeast Asian region. (UNESCO Bangkok, 2019).

Table 2: Participation in cross-country assessments in the Southeast Asia region, 2005-2019²

	Years conducted in EAP ³	Cambodia	Philippines	Lao PDR	Malaysia	Vietnam	Myanmar	Thailand	Indonesia	Timor-Leste	Singapore	Brunei - Darussalam
Early Grades Reading Assessment (EGRA)	Since 2015	✓	✓	✓		✓			✓	✓		
Programme for the Analysis of Education Systems (PASEC)	2011/2012	✓		✓		✓						
Progress in International Reading Literacy Study (PIRLS)	2006, 2011							✓	✓		✓	
Trends in International Mathematics and Science Study (TIMSS)	2005, 2011, 2015, 2019		✓		✓			✓	✓		✓	

² Loizillon, A., & Uys, P. (2020). *EAPRO Meta-analysis on what works to improve learning outcomes in primary education in East Asia and the Pacific. Literature Review: Learning Outcomes in East Asia and the Pacific (Volume 1)*.

³ East-Asia and the Pacific region

	Years conducted in EAP³	Cambodia	Philippines	Lao PDR	Malaysia	Vietnam	Myanmar	Thailand	Indonesia	Timor-Leste	Singapore	Brunei - Darussalam
Programme for International Student Assessment (PISA)	2003 2006 2009 2012 2015 2018		✓		✓	✓		✓	✓			
South-East Asia Primary Learning Metrics (SEA-PLM 1)	2019	✓	✓	✓	✓	✓	✓					✓

2.2 Object of the evaluation - SEA-PLM

SEA-PLM focuses on supporting ASEAN and SEAMEO member countries to better understand the status of student learning achievement, in order to improve the quality of their education systems (UNICEF, 2019). Assessing reading, writing, math and global citizenship, SEA-PLM is a high-quality learning assessment, with items and questionnaires contextualised to the regional context (UNICEF, 2019, p. 9).

SEA-PLM is nested within the regional Initiative for ASEAN Integration (IAI), which was launched in 2000 to provide regional integration in order to narrow the development gap within ASEAN. Five strategic areas for integration include food and agriculture, trade, micro, small and medium enterprises, education and health and well-being. The third IAI work plan states:

“Developing human capital is a major challenge for CLMV [Cambodia, Lao PDR, Myanmar and Vietnam] countries, with skills gaps widening as economies evolve. While access to basic education has increased, there remain large numbers of out-of-school children and quality is largely unaddressed.” (The ASEAN Secretariat, 2016, p. 8)

2.2.1 Emergence of SEA-PLM

SEA-PLM was launched as the world refocused education development goals away from access alone, and towards inclusive and equitable quality education for all. The idea informed by the progress of the global Learning Metrics Task Force (LMTF) initiative and was the brain-child of UNICEF EAPRO and SEAMEO in 2012 (UNICEF, 2019, p. 13). The concept was presented during the SEAMEO Council at the 25th SEAMEO High Officials Meeting in November 2012 and SEA-PLM was recommended for approval by the Council of Ministers. A Working Group was established in 2013, including UNICEF EAPRO, UNESCO Bangkok, UNESCO IS – Bangkok, ESCAP Statistics Division, SEAMEO Secretariat, SEAMEO INNOTECH and ACER (UNICEF, 2019, p. 13). The initiative was further conceptualised and endorsed at the 47th SEAMEO Council Conference in March 2013 and a cooperation agreement was established between UNICEF EAPRO and SEAMEO INNOTECH to coordinate the Regional Expert Reference Group (RERG), develop research papers to review education curriculums and assessment programmes across Southeast Asia, and to formulate strategic place for the future of SEA-PLM (UNICEF, 2019, p. 13). The RERG sat in late 2013 and included 30 curricula, assessment and examination experts from eight SEAMEO Member Countries, along with other partners (UNESCO Bangkok, UIS Bangkok, UNICEF and ACER). The early structure of SEA-PLM was shaped by this meeting.

At the 36th SEAMEO High Officials Meeting in February 2014, targeting grade 5 in the four domains was fully endorsed. The formal collaboration was then established between the SEAMEO Secretariat and UNICEF EAPRO and the agreement continued to 2018 and ACER was contracted under UNICEF EAPRO as the technical partner for SEA-PLM.

Brunei Darussalam, Cambodia, Lao PDR, Malaysia, Philippines and Thailand agreed to further support the process and a formal launching ceremony of SEA-PLM took place at the end of 2014 (UNICEF, 2019, p. 14). At the launching ceremony, the official Regional SEA-PLM Steering Committee was established, with membership of the initial Working Group being expanded to include several MOE officials from SEAMEO Member Countries and a revised Terms of References was established (UNICEF, 2019, p. 14). Consistent SEA-PLM endorsement took place at regional key High-Official meetings and updates on SEA-PLM were presented during SEAMEO High Officials Meetings (UNICEF, 2019, p. 14).

A range of visits and workshops took place including SEA-PLM Secretariat visits to several field trial countries in 2015, regional item development workshop in June 2015, a regional field operations and data management workshop in August 2015 and a regional workshop on coding in 2016.

Interest in SEA-PLM was raised within the region and at the end of 2015 Myanmar decided to join the field trial, followed by Vietnam, Philippines and Malaysia. Field trials were conducted in Brunei Darussalam, Lao PDR, Cambodia, Myanmar, Malaysia, Philippines and Vietnam between 2015 and 2018 (UNICEF, 2019, p. 15).

In 2016 SEA-PLM was included in the ASEAN Work Plan on Education (2016-2020), under Priority Area 2.2 *Improving the quality of basic education through quality-focused interventions* (The ASEAN Secretariat, 2016, p. 6).

2.2.2 Governance structure of SEA-PLM

Since inception, the SEA-PLM programme's structure has evolved. Originally, UNICEF EAPRO acted as the conduit between all parties within the partnership (SEAMES, ACER and UNICEF Country Offices) and facilitated all of the contractual arrangements for SEA-PLM (UNICEF, 2019, p. 4). UNICEF COs delivered advocacy for SEA-PLM with the national MOEs. They provided funding for implementation and ensured SEA-PLM queries were directed to the right people within MOE. UNICEF COs acted as the conduit between the MOE and all other partners and ensured continuous follow up with technical teams, highlighting any risks or challenges as they arose. UNICEF EAPRO defined the roadmap for SEA-PLM, identifying needs and ensuring relevant technical support. The regional office supervised all contracts under SEA-PLM, represented SEA-PLM when needed and appropriate, tracked progress, addressed risks and maintained continuous communication with UNICEF COs to deliver support and guidance. UNICEF EAPRO monitored and communicated regularly with technical suppliers to ensure progress and identify opportunities (UNICEF, 2019, p. 43).

Within this partnership, UNICEF identified the partnership with the SEAMES as crucial in positioning SEA-PLM within the regional intergovernmental governance structures and as a key to the success of SEA-PLM (UNICEF, 2019, p. 4). SEAMES delivered the regional mandate and facilitated the regional and national buy-in. This included coordinating regional workshops, steering committee meetings and processes of commitments. They also represented SEA-PLM at the global, regional and national levels and delivered policy expertise and guidance towards securing SEA-PLM as an integrated and sustainable programme.

The technical expertise and all direct country implementation support and training for SEA-PLM was provided by ACER, through a contractual commitment. ACER were engaged from 2013. External partners included NEQMAP, Brookings, UNESCO Institute of Statistics (UIS), INNOTECH and ESCAP.

As the original governance structure was designed for the inception period, in mid-2018 an external consultancy was undertaken to review, analyse and provide recommendations on how the governance mechanism for SEA-PLM can be improved and sustainable, ensuring quality founded within existing regional structures and aspirations. Consultation on the new governance structure was undertaken during the Regional SEA-PLM Consultation in 2018. As a result of the external consultancy, recommended changes included (1) that members of the SEA-PLM Steering Committee will only be Ministry of Education (MOE) country representatives, (2) that MOE representatives of non-participating SEAMEO countries should be invited to the Steering Committee on a no-vote basis, (3) that each MOE will have 2 representatives in the SC, and (4) that a Technical Reference Groups should be

established that would advise the Secretariat and Steering Committee (UNICEF, 2019, p. 46).

A Ministerial Declaration was pre-endorsed in draft form and was reviewed by countries during the Regional consultation in August 2018. A new approved SEA-PLM Secretariat structure that reports to its chairs, UNICEF and SEAMEO, and ultimately to the RRSC and the SEAMEO Council, was endorsed by countries in March 2019. It was agreed that the Secretariat is responsible for the day to day management of region and national technical and operational tasks in collaboration with countries and contractors, the SEA-PLM Secretariat is chaired and jointly overseen by UNICEF EAPRO and SEAMEO Secretariat and that the partnership between the two organisations is crucial for enabling the commitment of National Ministries of Education to implement SEA-PLM (SEA-PLM, 2019a). There was recognition that the structure of the Secretariat has shifted over the years based on available financial and human resources. It was decided that given its technical and political networks, the SEAMEO Secretariat would maintain the ultimate ownership and oversight of SEA-PLM through its internal structures with country membership and that SEA-PLM should be part of the country assessment governance and system. Under the new governance structure, SEAMEO contributes to technical development processes, strategic direction and operations of SEA-PLM. UNICEF contributes financial resources at regional country levels, technical expertise and operational leadership at the regional and country levels (SEA-PLM, 2019a, p. 6).

In 2019, the Technical Advisory Group (TAG) for SEA-PLM was established. The TAG seeks to provide independent advice on learning assessment matters and to play a consultative and advisory role on SEA-PLM methodology and results. The TAG chair and members are selected by the SEA-PLM Secretariat and mandated for one SEA-PLM cycle. The current TAG includes five experts, drawn from the Korean Institute for Curriculum and Evaluation; the University of Bath; the Chinese University of Hong Kong and Hong Kong Institute of Educational Research; Universiti Sains Malaysia; and Education Testing Service (ETS) Research Institute and the International Association for the Evaluation of Educational Achievement (IEA) (SEAMEO and UNICEF, 2019b).

2.2.3 Technical structure of SEA-PLM

An assessment framework has been developed and published for SEA-PLM, outlining the technical design of the assessment. The framework was developed in close consultation with national and regional experts and led by the SEAMEO Secretariat and UNICEF EAPRO. The SEA-PLM Assessment Framework was developed to identify '*what should be measured and in what way*'. Recognising the variations between countries, rather than being curriculum-based, SEA-PLM is curriculum-referenced and represents structures, conceptual underpinnings and overarching orientations across all ASEAN countries (ACER, SEAMEO & UNICEF, 2017).

The SEA-PLM Framework was reviewed in February 2015 by the Domain Technical Review Panels (DTRPs), established for each domain subject consisting of national experts in each domain (UNICEF, 2019, p. 14).

The assessment framework is designed for the first cycle of SEA-PLM implementation, aimed at students in Grade 5 or equivalent, but leaves scope for further development for SEA-PLM beyond the SEA-PLM 2019 first cycle. (UNICEF and SEAMEO, 2019a). For the future, it is suggested the framework could be adjusted to suit other grades should these be included in future rounds of SEA-PLM, and that the assessment programme could eventually provide measures of students' progress at key stages – middle primary school, towards the end of primary school; and towards the end of compulsory secondary school. (UNICEF and SEAMEO, 2019a, pp. 3-4)

The assessment framework includes four domains: mathematical literacy, reading literacy, writing literacy, and global citizenship. A separate assessment framework has been developed specifically for global citizenship. (UNICEF and SEAMEO, 2019a, p. 3)

SEA-PLM aims to measure both curricular and cross-curricular knowledge, skills and understanding that are likely to allow school-aged students to progress successfully through school and ultimately to play constructive and fulfilling roles as citizens in society. It adopts broad definitions for the domains of mathematics, reading, writing and global citizenship that are consistent with curriculum specifications but that allow for a focus on the extent to which students in a Southeast Asian context are able to make effective use of their knowledge in a variety of relevant contexts.' (UNESCO Asia and Pacific Regional Bureau for Education, 2018).

The programme focuses on comparisons in learning achievement both within and between countries. Programme documentation outlines that sub-national comparisons by social groups are important to inform decisions from an equity perspective, sub-national comparisons by school type are important to inform policy decisions, and cross-country comparisons are important to provide an external frame of reference (ACER, SEAMEO & UNICEF, 2017, p. 5 & 6). In addition, trends over time are purported to provide important information that contributes to monitoring movement towards the learning achievement goals. The trends monitored include changes in achievement at one grade level over time, and changes in differences between sub-populations over time (ACER, SEAMEO & UNICEF, 2017, p. 6).

2.3 Other partner support to improving access to quality education for all in the region

Within UNICEF's programme of support to education in the region, SEA-PLM falls under activity output 6.2.3: Better learning assessments and learning environments for children (EDU Area 2) (UNICEF EAPRO). A range of other bilateral and multilateral partners work within SEA-PLM participating countries to improve access to quality education for all.

Partners include the Global Partnership for Education (GPE), DFAT, the Embassy of Finland, UNESCO, the Department for International Development (DFID), the Asian Development Bank and the World Bank.

2.4 SEA PLM Logframe

The only logframe developed for the overall SEA-PLM programme was developed in July 2015, to support activities to December 2017. The stated plan was for the logframe to be further elaborated and also supported by a Theory of Change, developed collaboratively at a SEA-PLM Steering Committee. The accompanying narrative for the logframe states that the logframe will be used as a means of verification and will be regularly reviewed, validated and updated by the SEA-PLM Committee (SEA-PLM, 2015, p. 4). No further iterations, validations or verifications were identified during the desk review (See Annex F).

2.5 Reconstructed and refined Theory of Change (ToC)

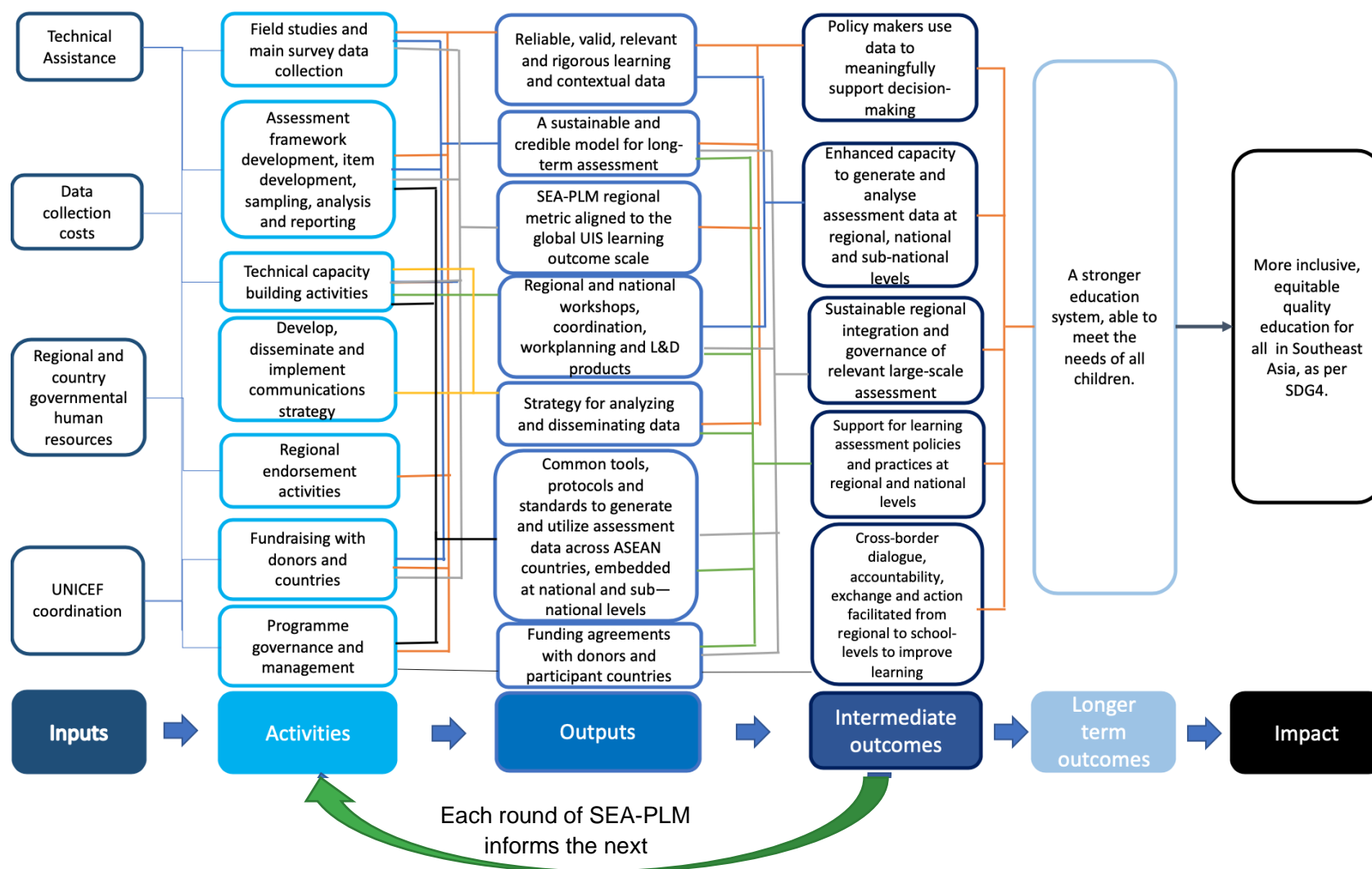
The reconstructed ToC was developed based on the literature review and desk study, as outlined in the ToR. The provisional reconstructed ToC was presented through a stakeholder workshop during the inception period and feedback from stakeholders was elicited. Based on the feedback the evaluation team refined the ToC. The refined ToC is what will be validated through the evaluation. The reconstructed and refined ToC will serve as the guiding structure for this evaluation.

The ToC for SEA-PLM based on the desk review has been implicit, rather than working to a formally prepared initial design. This is discussed in a paper prepared by staff of UNICEF, SEAMEO and ACER for the Oxford UKFIET conference in 2015 (Ahmad, 2015).

The implicit theory of how SEA-PLM programme implementation can lead to the desired outcomes and impact is rarely expressed in the language of a ToC in SEA-PLM documentation. While “impact” and “outputs” are referred to in the logframe, the language of the implicit ToC tends to focus on the programme vision, goals and aims.

One reference to a ToC map within the SEA-PLM 2019 Communication Strategy does not follow the usual logical chain of a ToC (inputs, activities, outputs, outcomes and impact) and rather maps goals to activities, to leavers of change, to systemic change and finally, impact (SEA-PLM SEAMEO UNICEF, 2019b, p. 18).

Figure 2 SEA-PLM Reconstructed ToC



2.5.1 Assumptions

Assumptions form the underlying beliefs about how a project will work, about the people involved in the programme and about the stakeholders. Assumptions help to explain the ToC process and the connections along the logical results chain. They help explain how and why proposed activities are expected to bring the desired changes about. Assumptions are often taken for granted and they may be based on opinions or beliefs.

Assumptions are often implicit but by stating them explicitly, the extent to which they hold can be assessed. Where assumptions do not hold, the causal chain is at risk and a new pathway to impact can be identified.

The following assumptions have been identified from inputs to activities, activities to outputs, outputs to intermediate outcomes, intermediate outcomes to longer term outcomes and longer-term outcomes to impact. These assumptions will be tested at the regional level and in the case study countries and weak points in the logical chain will be identified.

2.5.1.1 Inputs to activities

- Stakeholders have the opportunities (resources, time, conducive environment) to engage in SEA-PLM activities;
- Stakeholders have the motivation and incentives to participate in SEA-PLM activities; and
- The Secretariat has sufficient leverage within SEA countries to influence SEA-PLM participation and monitor SEA-PLM activities.

2.5.1.2 Activities to outputs

- Donors and participating countries have sufficient resources to invest in SEA-PLM and consider SEA-PLM Value for Money (VfM) and there remains sufficient resources from all sources, to continue to administer SEA-PLM activities
- Technical and capacity building inputs are sufficient to result in the desired outputs;
- There is sufficient common ground between countries to find agreement on the analysis and dissemination of SEA-PLM results;
- There is sufficient domain coverage, international and regional data and political will to support a common regional and global metric; and
- The data generated by countries is of sufficient quality to support a regional common learning metric.

2.5.1.3 Outputs to intermediate outcomes

- There is sufficient national capacity (technical capabilities, political will, resources) to analyse available data;
- Data is provided within the timescale needed to inform policy (within the policy window)
- Policy makers have sufficient capacity, incentives and political will and space to access and utilize data to inform decisions;
- The reach, fidelity and dose of capacity building activities are sufficient at national and regional levels;
- The costs of SEA-PLM are sustainable during integration into regional systems and beyond; and
- ASEAN countries see the value of a common approach to regional assessment and believe SEA-PLM is sustainable.

2.5.1.4 Intermediate outcomes to longer term outcomes

- The use of SEA-PLM data leads to the improvements of previous shortcomings in relation to relevance, equity and learning in the education system;
- Stakeholders have the capacity and incentives to solve education sector issues; and
- Stakeholders have the opportunities (resources, time and a conducive environment) to solve education sector issues.

2.5.1.5 Longer term outcomes to impact

- Changes in the education system positively affect relevant learning outcomes and equity.

3 Evaluation purpose

This section of the report outlines the purpose of the SEA-PLM evaluation and why it is being undertaken at this time. The primary and secondary audience and users, duty bearers and rights holders are described, and the objectives and scope of the evaluation are outlined.

3.1 Type of evaluation (why being done at this time)

The main purpose of this independent evaluation is to draw lessons learned and account for results to inform the next phase of the SEA-PLM Programme. The evaluation will be retrospective (*summative*) and cover (*secondary*) the conception period of SEA-PLM (2012-2015) and (*primarily*) the implementation of SEA-PLM 2019 from 2015 to 2019, but it will be forward-looking (*formative*) in providing conclusions and recommendations for regional- and country-level programming.

The first reporting cycle of SEA-PLM will be in 2020 and the dissemination of results will take place in 2020 and 2021 and throughout this time, SEA-PLM will move towards a longer-term Programme. For this reason, now is an appropriate time to explore the strengths and weaknesses of SEA-PLM governance and assess technical, managerial, funding and partnership perspectives (UNICEF, 2019).

3.2 Primary and secondary audiences

The primary audience of this evaluation is UNICEF EAPRO, SEAMEO, the SEA-PLM Secretariat, UNICEF Country Office Management teams and Education teams; the secondary audience includes governments, development partners, donors and external stakeholders such as contractors and other collaborators. The findings, conclusions, and recommendations will be used by both the regional and country offices to inform the development or implementation of their country Programme documents and adjust their strategies and implementation modalities as necessary.

3.3 Duty bearers and rights holders (and their involvement)

The right to education is a right of every human being and the principal right-holder is the person being educated (Monteiro, 2010). In the case of SEA-PLM, children in primary school in Southeast Asia are right holders and this is why the impact of the SEA-PLM programme must be their improved learning outcomes. The guarantee of the right to education is an obligation of State Parties in the instruments of International Human Rights Law. Therefore, it is the duty of governments to provide quality education for all. Parents and caregivers, communities, teachers, civil society organizations and the international community has a subsidiary and auxiliary responsibility to support states and individuals to achieve these duties (Monteiro, 2010). UNICEF and the international community can be a major participant in the realization of the right to education. This is captured in the Dakar Framework for Action, which introduces the role of the international community in the allocation of a larger share of resources to support basic education and ensuring education strategies complement other strategies for poverty elimination. The Paris Declaration on Aid Effectiveness supports rights-based approaches in terms of capacity building and accountability. In the case of SEA-PLM, the role of non-state actors include the provision of resources, supporting system strengthening and national ownership and advocating for human rights thinking and practice in aid delivery (UNICEF, 2007).

3.4 Evaluation objectives

The objectives of this evaluation are to (UNICEF, 2019):

1. Review the relevance, effectiveness, efficiency, impact, the sustainability of the SEA-PLM Programme and assess the extent to which it has been implemented in line with the objectives and global good practices on regional learning assessments;
2. At regional and national level, identify what strategic approaches, programmatic components and intervention modalities can help better position UNICEF and collaborators for SEA-PLM Programme and much broader in a fast-changing EAP region and global initiatives and in support of country programming in achieving the Sustainable Development Goal (SDG) agenda.

This evaluation will also strengthen and update the knowledge and evidence base on SEA-PLM in complementarity with other initiatives recently completed, on-going or recently launched by UNICEF in the region such as the literature review of what works to improve learning in the region. The evaluation will inform the planning and development of the second round of SEA-PLM, which will begin in 2021.

3.5 Scope of the evaluation

The time boundaries of the evaluation are 2012 to 2020, with a focus on the current implementation period of SEA-PLM (2015-2020).

The scope of the evaluation in terms of SEA-PLM organisation includes UNICEF country offices covered by SEA-PLM, UNICEF regional office, the Regional Secretariat, SEA-PLM participating countries and consultation with non-participating countries.

Geographically, the evaluation will cover the six Southeast Asian countries participating in SEA-PLM but will use three of these countries as case studies.

4 The evaluation framework

This section of the report describes the findings of the literature review undertaken during the inception period, any adaptations of the initiation ToR, the evaluation criteria and evaluation questions and the hypothesis to be tested within the evaluation. The enablers and barriers identified in this literature review are reflected in the Evaluation Matrix. Cross-references are included in the literature review summary below.

4.1 Summary of the literature review

Large scale learning assessments are estimated to be conducted in around 70% of countries globally (Lietz and Tobin 2016). These include national, regional and international assessments. International and regional learning assessments have many purposes, including comparing achievement between countries, identifying the determinants of national achievement and identifying factors that affect differences between countries. This literature review focused on three major international and regional learning assessments: PISA, SACMEQ and PASEC, drawing out lessons under two main headings:

- What impacts have been achieved by the learning assessments?
- What are the enablers and barriers in achieving those impacts?

Impacts achieved by the learning assessments

PISA is regarded by many as the gold standard of educational assessments (Breakspear 2012). Its results, published every three years, receive considerable attention globally, including in major media outlets, and have been known to hold influence over senior government officials and lead to radical changes in education systems. For countries that have performed well, the PISA results can serve to reinforce existing policies and may be co-opted for political agendas. In countries that perform less well or worse than expected, the effect has been described as a “PISA shock” with large-scale reforms implemented as a result (Baird et al 2011; Breakspear 2014). One such example was Germany in the first PISA assessment, which went on to enact widespread reforms, including generating national standards, reviewing curriculum standards to include PISA competencies and providing support for disadvantaged children (Breakspear 2012). The considerable influence that PISA now holds is seen in the inclusion of PISA metrics in national frameworks, plans and targets (Baird et al 2011; Breakspear 2014), however this has also received criticism (Breakspear 2014, Stanley 2013). The introduction of PISA rankings has been likened to an international sporting event (Stanley 2013). Critics highlight that the focus on rankings has been at the expense of nuanced sub-national evidence and efforts to enhance the intrinsic value of teaching and learning. In addition, with PISA only testing 15-year olds, some scholars have questioned the appropriateness of using PISA as the basis for reforms across the entire education system (Breakspear 2014).

SACMEQ is an example of a regional learning assessment which now brings together 16 ministries of education in southern and eastern Africa. The assessments test the reading and mathematics skills of Grade 6 pupils. Contextual information on the conditions that might influence teaching and learning is also collected. Combined, the assessments provide member countries with insights into the quality of education that their systems offer from an input, process and outcome perspective, allowing comparison cross-nationally and over time. Many and varied examples are provided in the literature of the influence that the SACMEQ has made since the first results were published in 1999.

For most member countries, SACMEQ has provided a significant step up in the quality of the data available for monitoring their education systems (Murimba 2005). It has also had knock-on effects for the quality of other data collected by governments, such as EMIS data (Murimba 2005). In terms of influences on policy making, the list of examples is long. It includes the formation of presidential commissions and national policy reviews as well as significant reallocations of funding for the construction of classrooms and the provision of classroom furniture and teaching materials (Howie and Plomp 2005). There are examples of the SACMEQ results uncovering practices and problems that were previously unrecognised, such as high repetition rates and the prevalence of over-age pupils who tend to leave school prematurely (Murimba 2005). SACMEQ has resulted in the establishment of new institutions, such as the Gender Unit in Uganda (HEART 2016), and new national monitoring systems, such as the National Assessment Project in Zambia (Howie and Plomp 2005). One further impact achieved by SACMEQ has been in terms of building the capacity within ministries of education (Murimba 2005). Staff have received training across the range of activities needed for rigorous evidence generation. They have also been instructed in the dissemination of findings and facilitating dialogue with policymakers and wider stakeholders.

PASEC is an international learning assessment that measures mathematics and reading levels among primary pupils in grades 2 and 6. Since its creation in 1991 by the Conference of Ministers of Education of Francophone Countries (CONFEMEN), assessments have been conducted in over 20 countries in Africa and Southeast Asia (Hounkpodote 2018). Unlike PISA and SACMEQ, PASEC's goal is not to provide a comparison of student attainment levels across countries; rather its purpose is to understand the factors which underpin and influence education quality (Bernard 2006).

In Senegal, for example, one of the objectives of the assessment was to gain insights into the effects of grade repetition on learning. It generated evidence that grade repetition is not associated with significantly better learning outcomes, contrary to the view held by three quarters of teachers. Based on the PASEC findings, a new policy was introduced prohibiting grade repetition in primary schools, while messaging in the media and seminars with teachers and school inspectors were implemented to explain the new policy. Interestingly, similar evidence collected in other countries in the region was less effective in influencing policy, due to political instability in Cote d'Ivoire and insufficient dissemination activities in Burkina Faso (Bernard 2006).

In Guinea and Togo, the PASEC assessment was used to explore whether the policy to recruit teachers with minimal or no formal teacher training was negatively affecting the quality of education delivered in schools. Controversially, the assessments found that these new teachers were not reducing the education quality – a finding which was welcomed by the governments, but not by teacher unions. The PASEC findings revealed that the effect of the reduced teacher training was countered by a higher average educational attainment among the new teachers and also their motivation to retain their contract. It also showed that, while the new teachers had received less training, that training had benefitted from being focused on practical teaching practice. In Guinea, the results led to the continuation of the teacher training programme with funding from the World Bank. In Togo, lower entry requirements were set for teachers, coupled with mandatory training (Bernard 2006).

Enablers and barriers

There are many examples and lessons to be learned from the international and regional learning assessments and the different the ways in which they have been implemented both internationally and within individual countries. The review presents these as lists of enablers and

barriers in order to assist policy makers, planners, assessment experts engaged in the planning and implementation of assessment frameworks.

At the macro-level, securing the strategic engagement of government and of key stakeholders within government is critical for success (Judgement criteria 1.1). Political instability can result in lost opportunities, however political change or wider educational reforms can create fertile soil for shaping new policies (Indicator 3.2 and 3.6). If the host ministry has a culture of learning (Judgement criteria 3.1 and Indicator 8.2) and if planning functions and processes centrally, experience shows that assessment results will have greater policy impacts. Extending involvement beyond ministry staff and academia, to include media, religious organisations, private sector, NGOs, parliaments, teachers' unions, parent groups and schools is also important.

The technical, advocacy and leadership skills and energies of national research coordinators and other leading individuals in the national agency responsible for the assessment are critical (Indicator 3.6, Judgement criteria 9.2). These individuals can set and maintain high technical standards and help to ensure the buy-in of results (Indicator 3.6). However, individuals and agencies specialised in learning assessments and research may lack other skills, such as advocacy, social mobilisation and capacity development (Judgement criteria 3.2, 9.1 and 9.2), that are needed to maximise the opportunities for educational reform. To maximise impact, partnerships can be formed with organisations with these strengths (Judgement criteria 1.1 and Indicator 1.2).

The involvement of international experts and the high technical standards and specifications of international and regional learning assessments give policymakers confidence in the data and findings produced (Judgement criteria 12.2). To increase impact, the assessment needs to be contextually appropriate for individual countries' policy contexts and underlying factors studied to help policy makers make the right choices (Judgement criteria 2.1). Participation in international and regional learning assessments involves large commitments in terms of human, material and financial resources, particularly for developing countries (Judgement criteria 1.2 and Indicators 9.1 and 9.4). Many countries now have a choice of assessments – national, citizen-led or regional/international, including a choice of different international assessments. Regional and international assessments are less well suited to tracking national standards and barriers than alternative options and they tend to require long time periods from planning and design to the publication of findings.

To be of use and achieve impact, the data needs to accurately represent what students achieve, it must permit valid comparisons between countries and contextual information is needed in order to interpret the findings. The information must serve the practical needs of policy makers and address their own policy problems (Indicators 1.4 and 3.2). To serve teachers, follow-up diagnostic assessments can be carried out to pinpoint specific gaps in teaching and learning. There are also dangers inherent to international and regional learning assessments. Cross-national studies can be criticised for unfair comparisons of national performance. Assessments tend to focus on one or two cohorts, but they are often used as the basis for reforms across the whole education sector. Also, it can take two or more years for assessment results to be published, so that they may be outdated and not reflect more recent policies.

The wide dissemination and open discussion of results is critical for maximising the impact of assessments. Allocating sufficient resources to this stage is important. Best results require that countries play a strong role in developing the strategy for dissemination (Indicator 3.4). Ministers and senior officials require specific briefings. Dissemination is best through multiple mediums with translations into national languages and different resources and approaches

targeted at different stakeholder groups. If key stakeholders do not attend dissemination events or if materials cannot be read because of language barriers, findings will be questioned and there will be a gap in understanding the reasons for any reforms enacted.

There are many examples of assessment results leading to 'policy borrowing' between countries. Good practice is for governments to conduct their own reviews based on the findings to determine the way forward. Teachers also will need support and targeted resources in order to help them to use and apply the assessment results. As with the dissemination phase, sufficient resources will be needed in order to act on the findings, particularly in developing countries, and bilateral and multilateral donors have a role to play here.

4.2 Any adaptations from the initial ToR

The evaluation is designed based on the Terms of Reference, informed by our technical proposal and discussions with UNICEF. Variations from the initial Terms of Reference are as follow.

1. The evaluation will include case studies of three countries which have participated in SEA-PLM, rather than of six countries as stated in the Terms of Reference.
2. All or some of the tasks originally envisaged as involving travel and face-to-face meetings by the evaluation team, potentially including the country case studies and central meetings, workshops, presentations and consultations, will be conducted remotely, by telecommunications (See Section 7).
3. The timing and sequence of activities and deliverables is adjusted as indicated in the Inception Report, and as may be made in agreement with UNICEF during the course of the assignment, rather than fixed against the tentative deadlines indicated in the initial Terms of Reference.
4. Deliverables will be submitted in well-edited form, consistent with the UNICEF Style Book 2019, as indicated in the Terms of Reference. Deliverable documents will be submitted in formats suitable for UNICEF to package and re-present under UNICEF livery and branding of the UNICEF Brand Toolkit 2012 where desired by UNICEF, rather than branded and packaged in those forms.

4.3 Evaluation criteria and questions

The evaluation will answer the following evaluation questions, as listed by Development Evaluation Criteria⁴ (DAC) are outlined in the ToR.

Relevance

1. To what extent is SEA-PLM aligned to i) the national development policies and priorities of national education stakeholders and ii) UNICEF's national, regional and global objectives and intended impacts?
2. What are the programming gaps or unaddressed needs? What could be done better?

⁴ The OECD/DAC definition of evaluation has been adopted by UNICEF and all major development agencies internationally. The definition contains five evaluation criteria that should be used in assessing development interventions: relevance, efficiency, effectiveness, impact and sustainability.' These are general criteria that should be used as a basis for developing evaluative questions through the full range of evaluations topics, i.e. from single intervention through to thematic, and ways of conducting the evaluation, e.g. joint evaluation. Taken together, these five criteria should provide the decision-maker with the essential information and clues to understand the situation and determine what should be done next.

Effectiveness

3. To what extent have the objectives and expected outcomes of SEA-PLM been achieved or are likely to be achieved?
4. What have been the major factors influencing the achievement or non-achievement of SEA-PLM objectives and activities? What were the enabling factors, barriers and bottlenecks?
5. What can UNICEF and its partners do to ensure the objectives of SEA-PLM are met in the future? What kind of initiatives should UNICEF prioritize at the country and regional?

Efficiency

6. What were the strengths and weaknesses of SEA-PLM management processes? How could management of SEA-PLM activities be improved?

Likely Impact

7. How has SEA-PLM contributed to national education systems and assessment practices and policies and discourses thus far? And at the regional level, what positive and negative changes has SEA-PLM brought about?

Sustainability

8. To what extent can SEA-PLM activities, plans and strategies be fully integrated and implemented by the government (s), both technically and financially? To what extent are they likely to continue?
9. What are the key barriers and bottlenecks towards achieving sustainability of SEA-PLM activities?
10. How can SEA-PLM attract other countries in Southeast Asia and better link other international and regional initiatives?

Equity and gender equality

11. To what extent is SEA-PLM conducive to supporting the most marginalized populations and genders (including those furthest left-behind)?

4.3.1 Prioritization of the evaluation questions

As part of the inception workshop with the Evaluation Reference Group, stakeholders were asked to select priority evaluation questions for each of the DAC, where more than one question is listed. Participants were asked to select which of the questions (under each criteria) is most useful to inform SEA-PLM. Clear choices for prioritization emerged for most of the criteria.

Relevance: Overwhelmingly (10 of 11 participants) felt it was more important to identify programming gaps or unaddressed needs and what could be done better, over assessing the alignment of national development policies and priorities and UNICEF objectives to SEA-PLM.

Effectiveness: Identifying the major factors influencing the achievement or non-achievement of SEA-PLM objectives and activities (and enabling factors and barriers) was seen as more useful to inform SEA-PLM than identifying the extent to which objectives and outcomes have been achieved or identifying what UNICEF and partners can do to ensure the objectives of SEA-PLM are met in the future (9:1:1 participants respectively).

Likely Impact: Participants were asked if SEA-PLM is too early in the stages of implementation for the question of 'likely impact' to be able to be answered. Most participants did not believe it is too early to answer this question (8 no and 3 yes).

Sustainability: Participants believed the extent to which SEA-PLM activities, plans and strategies can be fully integrated and implemented by the government and identifying the key barriers and bottlenecks towards achieving sustainability were equally important (5:5). No participants felt understanding how SEA-PLM can attract other countries to participate in SEA-PLM and better link to other international and regional initiatives was the most important sustainability question.

Following engagement with regional and country stakeholders, the evaluation questions were prioritised. It was agreed that impact and efficiency questions would be de-prioritised and will be answered where possible, but will not be the main focus of the evaluation.

4.4 Hypothesis to be tested

The hypothesis to be tested in this evaluation is the ToC. Based on the reconstructed ToC the overarching hypothesis to be tested through the evaluation is:

“The combination of SEA-PLM inputs and activities and desired outcomes are relevant to participating national stakeholders, are likely to meet the objectives of SEA-PLM, is financially and technically sustainable and is conducive to supporting the most marginalized populations.”

The evaluation matrix turns the ToC of the SEA-PLM into a set of hypotheses, including judgement criteria, indicators, data sources and analytical approaches, on which the evaluation design and instrumentation is based. This is explained further in 5.2. The evaluation matrix can be found in Annex E.

5 Methodology

This section of the report outlines the conceptual framework, the evaluation methodology, the country selection for case studies, data sources, data collection methods, data collection instruments, data analysis techniques, equity and gender considerations and ethics guiding the evaluation.

5.1 Conceptual framework

We propose to undertake a non-experimental, theory-based Process Evaluation to answer the evaluation questions. Therefore, the principal organising structure and framework for the evaluation is the ToC, however we will follow the key functions of process evaluation.

Process Evaluation does not make use of a comparison group, seeking to make causal inferences, but rather seeks to make credible claims about the SEA-PLM programme’s progress along the inputs, activities, outputs and initial outcomes of the ToC. This is important during the early stages of an intervention, to inform the final implementation model.

The aim of the process evaluation is to explain how SEA-PLM works and will examine the processes through which the programme generates outcomes in different contexts. The figure below illustrates schematically key functions of process evaluation.

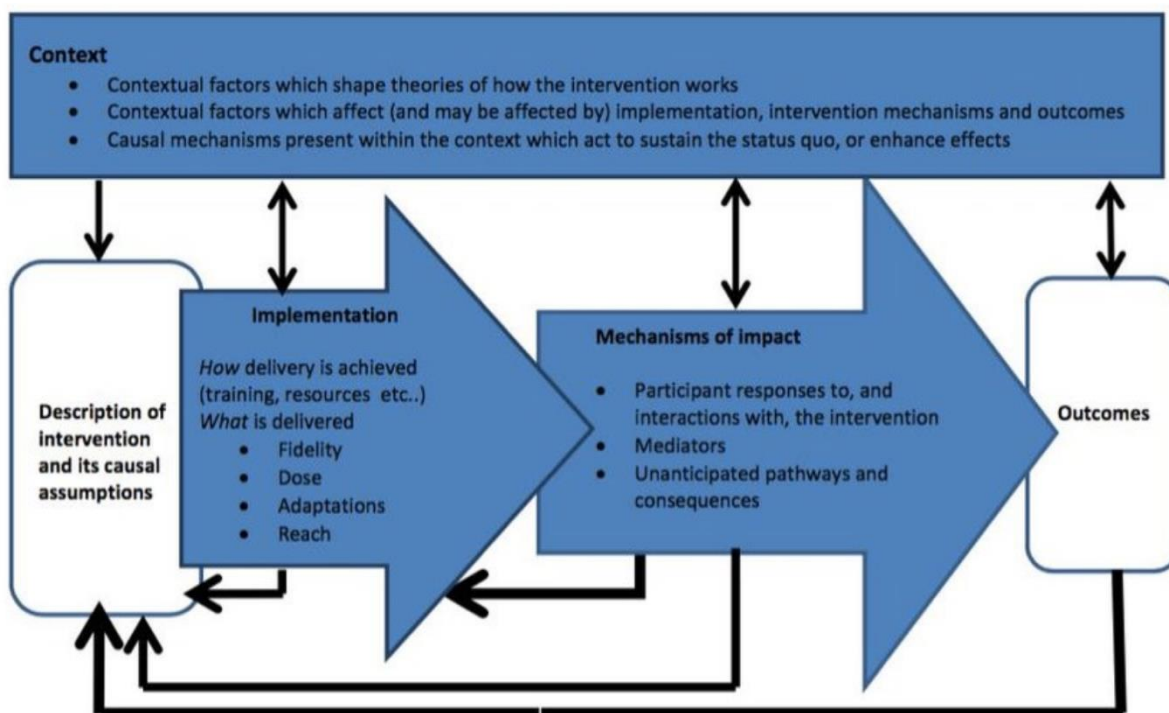


Figure 32 Key functions of process evaluation, from Moore et al, 2014

5.2 Evaluation approach

The evaluation will explore the interactions between context, SEA-PLM design and causal assumptions, how delivery was achieved, how participants responded to and interacted with SEA-PLM and will assess the validity of the results chain, as described in the ToC.

There are three components of the evaluation methodology which enable the explanation of how SEA-PLM works and the examination of the processes through which the programme generates outcomes in different context. They are:

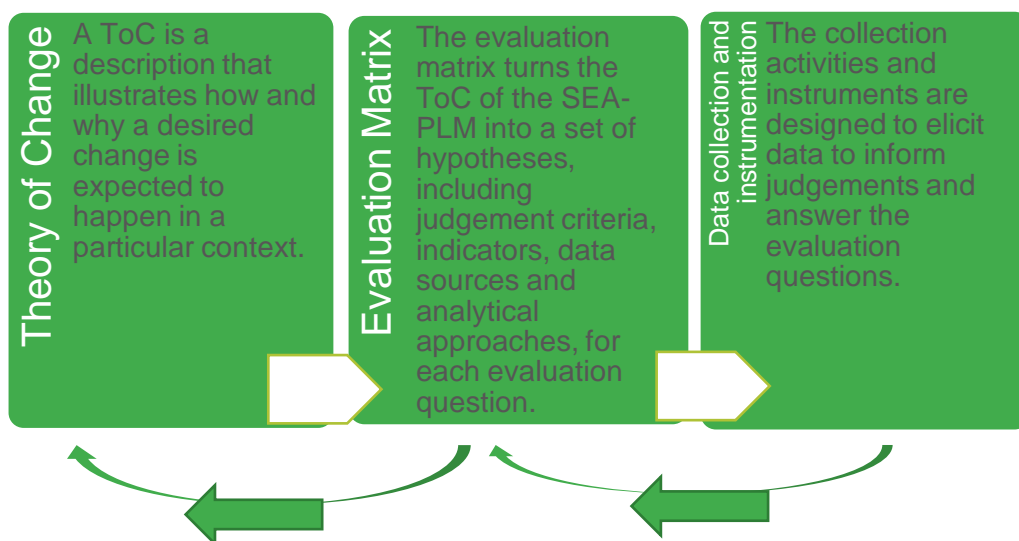
1. The ToC;
2. The Evaluation Matrix; and
3. Data collection instrumentation.

The evaluation methodology begins and ends with the ToC. The ToC describes and illustrates how and why a desired change is expected to happen in a particular context. The Evaluation Matrix turns the ToC into a set of judgement criteria, indicators, data sources and analytical approaches for each evaluation question. Each aspect of the ToC and the assumptions of the ToC are included in the Evaluation Matrix.

The data collection activities and instruments are designed to elicit data to inform judgements and to answer the evaluation questions.

The extent to which secondary data sources can be used to develop indicators has been assessed and the questions for primary data collection have been developed based on filling any remaining gaps.






Figure 4 The three components underpinning the evaluation methodology



The data produced is then analysed based on the techniques outlined in the Evaluation Matrix to become indicators and judgements are made based on the indicators based on the judgement criteria explicated in the Evaluation Matrix. These judgements are brought to bear on the ToC and provide answers to the evaluation questions.

This process is outlined in Table 1, using evaluation question one as an example. The Evaluation Matrix is outlined in full in Annex E.

Table 3 Example of how the evaluation matrix will be used, with reference to evaluation question one

Evaluation Criteria	Evaluation question	Judgement criteria	Indicators	Research methodology	Source of information	Link to the ToC
Relevance	To what extent is SEA-PLM aligned to i) the national development policies and priorities of national education stakeholders and ii) UNICEF's national, regional and global objectives and intended impacts?	1.1 Evidence of national development policies and priorities alignment with SEA-PLM objectives 1.2 Evidence of UNICEF global, national and regional objectives aligning with SEA-PLM intended impacts	National stakeholder support for SEA-PLM Country and national institutional needs against programme delivery UNICEF policy and programme documentation aligns with SEA-PLM ToC There is sufficient domain coverage, international and regional data and political will to support a common regional and global metric	Desk research, Qualitative Comparative Analysis (QCA)	Country policies, UNICEF national and regional and global documentation Programme documentation Key Informant Interviews (KII)	Context, inputs and activities
		 <p>How will we judge if SEA-PLM is aligned to national and UNICEF objectives?</p>	 <p>What will be the indicators we will need to make these judgements?</p>	 <p>What techniques will we use to develop these indicators?</p>	 <p>Where will we get the data from to apply these techniques?</p>	 <p>Which parts of the TOC does this evaluation question relate to?</p>

Using the Evaluation Matrix, the evaluation team will:

1. Identify barriers to delivery, uptake, individual change and institutional reform.
2. Review how delivery is achieved, by (1) identifying how the programme has been established, including institutional factors, partnership factors and resourcing; (2) Identifying what contextual factors affect implementation; (3) Identifying any adaptations that might be required to improve implementation, take up and individual and institutional change; and (4) Checking the fidelity of the programme against the ToC and programme documentation⁵.
3. Identify the likely mechanisms of impact⁶, including exploring country responses to, and interactions with SEA-PLM and exploring any unanticipated pathways and consequences of the programme.
4. Identify any contextual issues that shape theories of how the intervention works, contextual factors which affect (and may be affected by) implementation, intervention mechanisms and outcomes and any causal mechanisms present within the context which act to sustain the status quo or enhance effects and the mechanisms for impact.

This will be done by empirically verifying activities, outputs, and outcomes and assumptions posited along the causal chains in the ToC and draw conclusions about what is working and what is not working in SEA-PLM and likely impact.

5.3 Data

In this section of the report, we outline the data collection methods, data collection instruments and our approach to data collection consistency and quality.

5.3.1 Data collection methods

5.3.1.1 Inception report development and Desk Research

The process undertaken to develop this inception report has been outlined in Section 1.2. This report has been developed to outline the key approach, methodology and workplan for the evaluation. It includes a reconstructed ToC which has been developed based on SEA-PLM programme documentation and an evaluation matrix based on the final evaluation questions, and outlines the judgement criteria, indicators, data sources, and links to the ToC.

The ToC has been developed to inform the evaluation and provide the evaluation team and programme participants with a common language regarding how change is theorised to take place through the programme. An inception workshop took place during the inception period and included validation of the ToC.

As part of the inception period, the team reviewed available programme materials including strategic planning documents, any programme logic models, implementation work plans, and available data from the programme. This review has informed the availability and quality of data available to answer the evaluation questions and gaps to be filled through qualitative and quantitative data collection.

⁵ Fidelity evaluation is the “extent to which delivery of an intervention adheres to the protocols and program model originally developed” (Mobray, Holter and Teague, 2003)

⁶ It is recognised that as the programme has not completed a full cycle yet, impact of student learning cannot be assessed. However, the likely mechanisms for impact can be explored by testing the assumptions that exist along the results chain.

In addition, a review of literature on international and regional assessments has been undertaken during the inception period. This includes a review of the types and history of large-scale assessments, evidence of the impact of assessments and the enablers and barriers to achieving impact.

5.3.1.2 Qualitative data collection

Our evaluation methods are predominantly qualitative in nature, allowing us to understand what change is occurring, how and why. Qualitative data will be collected from UNICEF, regional governance actors, national government actors, ACER and the TAG. The project has a focus on capacity development, which is not readily quantifiable and many of the evaluation questions are 'how' and 'why' questions (rather than 'how much'). The purpose of the qualitative analysis will be to enable the evaluation team to answer evaluation questions related to participation barriers, the extent to which individual and institutional capacity interacts with achieving programme outcomes and the validity of the ToC in different settings. The main focus will be to collect data along the results chain from inputs, activities, outputs to outcomes, and identify any weak links in the ToC.

The desk-review of SEA-PLM documentation will inform the identification of key stakeholders at the regional and country levels for interview. The desk-review is the basis of the reconstructed ToC and in combination with the ToC will allow the evaluation team to identify key actors engaged in SEA-PLM activities and key decision makers at the regional and country levels.

A country case study approach will be utilized, and qualitative data collection will be conducted in three of the six participating countries through virtual or face-to-face country visits.⁷ Consultations with countries other than those featured as case studies (including non-participating countries) will be light touch and limited to a conversation with one government respondent and one UNICEF focus group discussion for each country. The purpose of these interviews is to situate the SEA-PLM experience of the three case studies within the wider context of the five SEAMEO countries yet to participate and the three participating countries with which case studies are not conducted.

5.3.1.3 Quantitative data collection

An online survey will be administered across participating countries in order to assess the extent to which experiences with SEA-PLM are heterogeneous. The instrument will follow qualitative case study data collection and will be constructed based on the findings of the three country case studies. The survey will be disseminated to as many actors as possible. During the qualitative data collection period, respondents will be asked to share names and email addresses for the wider set of actors engaged in SEA-PLM or with a stake in SEA-PLM. Invitations and participate in the surveys will be sent to all of these people.

5.3.1.4 Sampling for Country Case Studies

The selection of case study sites is informed by consideration of a theorised sampling approach that will produce data from a diverse set of countries that are collectively likely to produce data on a range of experiences to inform future decisions.

Sampling '*is a core design issue because the purposeful sample will determine what you learn about*' (Patton 2015). Through a carefully chosen sampling strategy, the most relevant sample can be selected, and rigour can be ensured. The chosen approach to sampling for the country case studies is designed to generate responses from small numbers of individuals and groups

⁷ See Section 7 on Risk Management and COVID related scenarios

that are representative (though not statistically) of groups relevant to SEA-PLM and which allow some identification of heterogeneous contributions and experiences.

Purposive sampling can serve this purpose. Purposive sampling involves selecting a sample of 'typical' and 'extreme' cases based on available data. In essence, the aim is to select countries that cover the range of national and education system features.

Therefore, our proposed selection for the three case studies is:

1. Lao PDR;
2. Cambodia; and
3. Malaysia.

Lao PDR represents a small education system with relatively less experience in large scale assessments. Cambodia is a larger system, with somewhat more established experience and Malaysia is a country with significant previous participation in international surveys.

Additionally, as has been outlined in Section 2 of this report, Cambodia has seen a deterioration in many education indicators, Laos has had some very positive progress but remains challenged in some areas and Malaysia has shown progress and very little deterioration across the range of education indicators. This sample also represents the countries with the highest (Cambodia) and the lowest (Malaysia) pupil-teacher ratios of the SEA-PLM participating countries, the highest (Cambodia) and the lowest (Lao) primary school life expectancy, the highest (Malaysia) and the lowest (Cambodia) survival rates to the last grade of primary education and the highest (Cambodia) and lowest (Malaysia) percentages of enrolled overage children.

We propose that the country for the first case study is Malaysia. This will also serve as a pilot for the other two. That first case study will be undertaken by our Team Leader and evaluation methodology lead.

5.3.2 Data collection instruments

The qualitative research tools will consist of Key Informant Interviews. Semi-structured questionnaires will be designed in alignment to the evaluation questions, will be informed by the evaluation matrix and will be conducted with a broad variety of stakeholders. Instrumentation will be tailored according to the category of respondent to ensure the most relevant and valuable insights are obtained.

Instruments will follow the results chain from inputs to outcomes but will use the language of SEA-PLM activities and SEA-PLM expected outcomes rather than the language of evaluators. Instruments will begin with requesting information on the informant's role and previous roles in order to ascertain early in the interview where the participant's input will be most valuable to the evaluation. Evaluators will then focus on those questions.

UNICEF evaluation standards state that right holders (children, parents, teachers in this case) should be engaged during evaluations. However, SEA-PLM is not currently engaging with schools and communities directly and is rather focused on engaging governmental stakeholders predominantly found in capital cities. For this reason, the evaluation will not engage rights holders.

Data collection tools can be found in Annex H.

5.3.3 Data collection consistency and quality

Data collection instruments will be standardised across the three countries, by sets of informants.

The first country study will act as a pilot to inform tool finalisation. After each interview, the evaluator will reflect informally on the tools and process and identify areas for improvement. The data collection tools will thus be refined iteratively within the first of the case study visits.

UNICEF will support introductions to case study country actors. UNICEF national teams will work with the evaluation team to determine which stakeholders should be approached for interview and a snowball approach will be taken throughout the interviewing process.

5.4 Data analysis

The evaluation team will use qualitative comparative analysis and synthesis techniques, carefully ensuring non-bias using triangulation and informed judgements.

5.4.1 Qualitative Comparative Analysis

Within the structure of the overarching ToC, the evaluation will use elements of Qualitative Comparative Analysis (QCA). QCA is a useful method to identify different factors or combinations of factors that are likely to have led to a specific outcome, in a given context.⁸

The qualitative analysis will allow the evaluation team to identify the underlying assumptions related to the ToC and the extent to which these did or did not hold in the different contexts.

5.4.2 Synthesis

The analysis will use applied case study comparative analysis, whereby each country will be considered a 'case' within the evaluation. The similarities, differences and patterns across the cases will produce knowledge that enables generalisation about causal questions along the results chain, across the diverse SEA-PLM contexts. The evaluation matrix and the ToC components (inputs, activities, outputs, outcomes and the assumptions that underpin them) will form the analytic constructs to be compared across cases.

To ensure efficient analysis of the data, we will keep a running record of evidence against each evaluation question in an evidence matrix in an Excel document. The Desk Research outputs will be reviewed to populate the evidence matrix. In the case of the country case studies, immediately after each interview, the interviewer will add new evidence emerging from the interview to the evidence matrix. The evidence matrix will be reviewed and discussed by the full team regularly, with a particular focus on points where new or conflicting evidence is emerging that needs to be explored further. In some cases, a need for a short follow-up conversation or email with a respondent may be identified to clarify or follow-up on a response.

5.4.3 Triangulation, bias and informed judgements

With qualitative interviews, it is often likely that respondents will provide biased answers. This is why a range of respondents is needed for triangulation. The data of each respondent will be organised around the themes / questions which then provides an overview of the frequency and strength of responses for each theme. Furthermore, data from interviews and document reviews is triangulated in order to provide nuanced answers to the evaluation questions. In the case of

⁸ See, for example: Baptist, C. and Befani, B. (2015): Qualitative Comparative Analysis: A Rigorous Qualitative Method for Assessing Impact. Available at: <http://www.coffey.com/assets/Ingenuity/Qualitative-Comparative-Analysis-June-2015.pdf>. See also: *Qualitative Comparative Analysis and the Study of Policy Processes*. Journal of Comparative Policy Analysis, Research and Practice Volume 19, 2017 - Issue 4: Special Issue: Validating methods for comparing public policy: Perspectives from academics and "pracademics".

contradictory and opposing views the evaluation team will represent these views in our analysis. Both for primary and secondary data, the evaluation matrix will help to structure the results and the findings.

In addition, when synthesising data, the evaluation team will be cognisant that SEA-PLM has been implemented only recently. So, it is likely to have involved activities and levels of resources and commitment associated with the introduction of an innovation, as distinct from its long-term continuation as an institutional programme. The data on the experience to date will thus not be reflective of what would apply in a longer-term continuation, in terms of resource, interest and commitment or technical capacity. Informed judgements will be made as part of the evaluation looking forward, which cannot be based simply on extrapolating experience to date looking backward.

A validation workshop will be held with the Reference Group to review preliminary findings and conclusions and to prioritise recommendations. From this workshop and other meetings, the draft report will be revised and finalised.

5.5 Equity and gender

Our approach to mainstreaming equity and gender into the evaluation is captured in the Evaluation Matrix (Annex E). In order to answer the evaluation question “*to what extent is SEA-PLM conducive to supporting the most marginalized populations and genders (including those left behind)?*” the evaluation team will seek evidence that equity and inclusion has been mainstreamed into programme activities and that the data SEA-PLM produces will be used to inform equity decisions. The evaluation team will check if assessment items are checked for differential item functioning (DIFF analysis) by gender and location and will explore if SEA-PLM data is able to capture inequality. Furthermore, the evaluation team will explore country capacity to disaggregate data by social groups and explore if SEA-PLM policy makers are considering equity, inclusion and gender as an essential component of their policy questions.

5.6 Ethics

The design and conduct of this evaluation is guided by Cambridge Education’s core values of integrity, respect and excellence as we constantly stretch our thinking to find extra value for our clients and their end-users. These values also guide our rigorous internal ethical standards, which are aligned with the United Nations policies and standards for ethics and evaluations. Applicable ethical norms and procedures of UNICEF and Cambridge Education will be applied throughout the evaluation.

The evaluation methodology will be consistent with the [UNICEF Evaluation Policy \(2018\)](#), [the United Nations Evaluation Group \(UNEG\) Norms and Standards for Evaluation \(2016\)](#), [UN SWAP Evaluation Performance Indicator](#), [UNEG Ethical Guidelines \(2008\)](#), [UNICEF Procedure for Ethical Standards and Research, Evaluation and Data Collection and Analysis \(2015\)](#), as specified in the Terms of Reference. The final evaluation report will conform to [the UNICEF-Adapted UNEG Evaluation Report Standards \(2017\)](#), including consideration of a gender equality perspective and human rights based approach, including child rights.

Photographs, audio and video recording may be used for gathering and analysis of data and for presentation and dissemination purposes. Permission for using any photographic, video or audio recording as part of the evaluation will be sought in advance. Before conducting any interviews or group discussions, participants will be made aware of the extent to which the information they provide is confidential and will be given the opportunity to withdraw at any point.

Cambridge Education requires all its staff, suppliers and implementing partners to comply with the highest standards of ethical behaviour in all respects and at all times, as reflected in Cambridge Education's mandatory adoption of its Ethics Policy and engagement in relevant training. All Cambridge Education evaluation team members involved have satisfied Cambridge Education's due diligence process and prior to commencing any fieldwork will have signed Cambridge Education's Ethics Policy Statement.

In addition, any potential conflicts of interest are taken into account when engaging staff, suppliers and implementing partners on any assignment and issues of payment and compensation are clearly laid out in all contracts and managed by CE's Commercial experts. Cambridge Education has zero tolerance around issues of fraud, bribery, corruption, child safeguarding and harassment; and where necessary will cooperate with the appropriate authorities to investigate alleged transgressions, including sanctions where required.

5.7 Limitations

There are four key limitations to the evaluation:

1. A limitation to the evaluation is the timeframe. Because the results of SEA-PLM 2019 will only be released in late 2020, the evaluation will not be able to assess impact of results on children and education systems. Rather it will take a more formative approach, looking at intermediate outputs and outcomes and likely impact.
2. Secondly, accessing sufficient comparable information across the set of implementing countries, and from the five other SEAMEO countries, may be difficult, given the diversity and range of these countries. However, a combination of our proposed approach and the assistance of UNICEF's EAPRO and Country Offices is expected to address this challenge satisfactorily. The Terms of Reference states that each of UNICEF's involved Country Offices will facilitate in these respects, by providing a designated focal point person to support the evaluation, providing all the information, resource documents and contacts necessary for the evaluation, facilitating communications with implementing partners in-country, organising and facilitating logistics, security, meetings and workshops, with financial support from UNICEF's regional assessment budget, and by providing comments on key deliverables to minimise factual errors, misinterpretations and omissions. Due to the current COVID-19 pandemic, UNICEF Country Offices are heavily engaged with supporting remote learning and back to school planning. This emphasises the need for clear and timely communication of needs by the Evaluation Team and effective follow-up by the regional office with the designated focal point in country offices to ensure deadlines are adhered to as much as possible.
3. Thirdly, a significant challenge is expected to be the limitations of availability of cost and resource data disaggregated from general government expenditure and hidden costs borne by various parties. Typically, countries do not cost their activities such as participating in surveys separately. For this reason and due to the highly labour-intensive nature of doing in-depth financial exegesis and analysis, adherence to the principals of VfM will be assessed in the evaluation, as there are not sufficient resources available for a cost-effectiveness component.
4. Finally, a serious limitation to the evaluation is the amount of resources allocated to the evaluation. For example, the country case studies have been allocated 5 days of evaluator time to conduct data collection and 5 days of evaluator time to analyse all case study data and write the report. This highlights the importance of keeping the scope of the evaluation clear and tight, focusing resources where they will provide the most added value and prioritising evaluation questions (as described in this inception report).

6 Workplan

It will be of the utmost importance that the team carries out its tasks and produces its working products to schedule, and that inputs and feedback from UNICEF and the Evaluation Reference Group are timely. It has been agreed that feedback from the Evaluation Reference Group will be consolidated through a single point of contact with the evaluation consultancy team. This is so that the process can proceed efficiently and in a timely manner.

Timing and maintaining forward progress through the stages of the assignment will require active attention. Work-planning will be undertaken and agreed between the consultancy team and UNICEF during the inception planning phase. The full workplan is presented below.

Figure 5 Workplan

Version 7
08-Oct-20

**WORKPLAN SEA-PLM evaluation
UNICEF EAPRO/Cambridge Education, 2020**

2020	JANUARY				FEBRUARY				MARCH				APRIL				MAY				JUNE				JULY				AUGUST				SEPTEMBER						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
	Wk	6	13	20	27	3	10	17	24	2	9	16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	31	7	14	21
1	Inception phase: DELIVERABLE 1																																						
1.1	Team mobilisation, familiarisation and preparatory discussions (remote)																																						
1.2	Team & UNICEF kick-off and inception planning meeting, including review of Theory of Change																																						
1.3	Literature search and review on large scale assessments																																						
1.4	Desk review of SEA-PLM documents																																						
1.5	Preparation of inception report, including evaluation theory of change, methodology, tools, limitations & mitigations, revised workplan, bibliography																																						
1.6	Presentation to SEAMEO and EAPRO																																						
1.7	Team and UNICEF meeting to discuss options to proceed in light of COVID-19 travel restrictions.																																						
D1	Submit inception report (a draft; b revised, c revision 2)																																						
2	Case study reports: DELIVERABLE 2																																						
2.1	Data collection relating to 11 SEAMEO countries																																						
2.2	Data analysis																																						
2.3	Case studies (3 countries) and preliminary findings																																						
D2	Submit case study reports (a draft; b revised)																																						
3	Draft evaluation report: DELIVERABLE 3																																						
3.1	Draft evaluation report																																						
3.2	Validation workshop - review preliminary findings																																						
3.3	Remote consultations to validate and revise the report																																						
D3	Submit draft evaluation report (a draft; b revised)																																						
4	Final evaluation report, brief, video and presentation: DELIVERABLE 4																																						
4.1	Finalise the evaluation report, all components																																						
4.2	Produce 4-page brief for general audience																																						
4.3	Produce PPT summary of final report																																						
4.4	Produce video version of PPT presentation																																						
4.5	Presentation to SEAMEO and EAPRO																																						
D4	Submit final evaluation report, brief, video and presentation (a draft; b revised)																																						

- Remote working
- Deliverable
- Comments from Client

Workplan and schedule, to be reviewed and developed. Specific dates to be agreed, and workplan managed and updated, by agreement between Cambridge Education and UNICEF during the assignment.

6.1 Evaluation management and quality assurance

Support to the day-to-day management of the evaluation, and guidance to the evaluation team will be provided by the EAPRO Evaluation consultant (EC) and the SEA PLM Project Manager (PLMPM). They will be responsible for the day-to-day oversight to ensure a smooth process and will also contribute to the dissemination of the evaluation once produced. Management of the evaluation budget rests with the Regional Education Advisor.

The EC and PLMPM will receive support from the EAP Regional Evaluation Advisor and the Regional Education Advisor and they hold decision making authority on the evaluation process. Further support to the evaluation process will be drawn from country office colleagues engaged in SEA-PLM. The Regional Evaluation Advisor has additional responsibility for ensuring the overall quality assurance for the evaluation and formal signing off on all evaluation deliverables.

This evaluation will also be guided by a reference group. The reference group will have the following roles.

- Generally, advise the evaluation management team on various aspects of the evaluation and help this team make decisions.
- Contribute to the preparation and design of the evaluation.
- Assist in identifying internal and external stakeholders to be consulted during the evaluation process.
- Participate in review meetings organized by the evaluation management team.
- Provide comments and substantive feedback from a technical point of view to ensure the quality of the second draft and final evaluation reports.
- Propose improvements/inputs to the preliminary recommendations.
- Play a key role in learning and knowledge sharing from the evaluation results.
- Contribute to disseminate the findings of the evaluation.

The quality of the evaluation team itself is the primary source of quality assurance for this evaluation. However, in addition to the core team, the Project Director from Cambridge Education, is ultimately accountable for evaluation deliverables and all work undertaken as part of the evaluation will be reviewed by the Project Director prior to delivery to UNICEF.

UNICEF will provide quality assurance on all evaluation tools and documents based on the [UNEG's](#) and [UNICEF's](#) norms, standards, and guidelines as well as on other best practices related to Programme evaluations. Once approved, the final evaluation report will be submitted to the UNICEF's global evaluation reports oversight system ([GEROS](#)) for an independent quality rating. The report and the review will be made available on the UNICEF Internet website, in compliance with the commitment for transparency of evaluation findings.

6.2 Roles and responsibilities

The evaluation team will be led by the Team Leader, Dr Rachel Outhred. The Team Leader is responsible for overseeing and managing other team members, managing communications with UNICEF, organizing key meetings and workshops and responding to comments from evaluation stakeholders. Rachel will conduct the first country case study and treat it as a pilot for the case study instrumentation. Rachel will write the country report for this country and will be responsible for synthesising across case study countries. Other evaluation team members include Jeffrey Marshall. Jeffrey will conduct two of the three country case studies, will write the country reports for these two countries and will support Rachel in synthesising across case

study countries. Jason Pennells role was to contribute to the Inception Report by preparing the desk study and the literature review.

Support to the evaluation team members will be provided by Cambridge Education's support team comprising of a Project Principal, Emma Mba, (responsible for quality assurance of all deliverables and technical advisor) and the Project Manager, Alice Wright, who will handle management and administration of contractual and invoicing matters with consultants and with UNICEF.

The role of the UNICEF country office Education and Evaluation teams will do the following:

- Designate a focal point for supporting the evaluation.
- Provide the evaluation team with all information, resource documents and contacts necessary for the evaluation.
- Facilitate the communication and coordination between the evaluation team and UNICEF's implementing partners in-country.
- Organise / facilitate logistics, security, meetings and workshops.
- Provide comments on the key deliverables to minimize factual errors, misinterpretations, and omissions.

Draft and final deliverables will be submitted to UNICEF.

UNICEF will establish and facilitate the Evaluation Reference Group.

6.3 Communication and coordination

The Team Leader is responsible for communications with UNICEF. In accordance with our internal procedures, Cambridge Education will maintain regular contact with UNICEF and provide backstopping to technically support the experts during the contract period, carry out quality control of the services, support the preparation of reports and, if needed, refocus and / or rebalance the experts' interventions in space and time.

UNICEF will liaise with UNICEF, regional and country stakeholders as appropriate, to review and comment on the draft and final deliverable submissions within agreed timeframes, and will gather, collate, consolidate and forward to the evaluation team the comments as a single batch on each occasion.

6.4 Logistics and support needed from UNICEF

The evaluation team also requests the following services from UNICEF's EAPRO (Bangkok) Office, and where relevant, from the participating UNICEF Country Offices.

1. Confirm and handle approval processes and manage the interface with other parties to maintain the smooth and efficient forward progress of the evaluation assignment and its reporting.
2. Provide space and access to facilities at the UNICEF office in Bangkok or other suitable venues, for meetings, workshops and presentations, in cases where these meetings are held face-to-face, where these may be required during the duration of the assignment.
3. In liaison with the Evaluation Team Leader and Cambridge Education Project Manager, set up remote teleconference meetings, and invite participants, where applicable for workshops, presentations and meetings held in lieu of face-to-face events.
4. Facilitate access to relevant UNICEF colleagues in the EAPRO office and in Country Offices and identify and facilitate contact with necessary interviewees for the Key

Informant Interviews for the case studies and for the light-touch consultations with countries not included in the case studies.

5. Collate feedback, review it for consistency against previous agreements and ensure conflicting views are resolved internally before providing consolidated feedback to the evaluation team from the Evaluation Reference Group, where this is required, and provide timely comments on the reports, instruments and plans as necessary to enable the assignment to keep to schedule.
6. Carry out, provide, or arrange with partners involved, any translation or interpretation services required in the course of this evaluation (for documents and meetings), or for wider dissemination of documents produced by the evaluation team. (All deliverables for this assignment will be submitted in English.)
7. Provide support letters and introductions for the evaluation team members to facilitate their work, including for remote contact, accessing information or for travel, as applicable.

6.5 Deliverables and reporting

Deliverables and payment schedule

No.	Deadline	Deliverable	Percent payment
Deliverable 1	October 2020	<p><u>Phase 1</u></p> <p>Inception Report and presentation, including:</p> <ul style="list-style-type: none"> • Literature review • Methodology and data collection and analysis tools including sampling strategy • Foreseen limitation and mitigation measures • Workplan with timeline • Annotated Bibliography <p>Presentation to SEAMEO and UNICEF EAPRO (remotely)</p>	20%
Deliverable 2	February-March 2021	<p><u>Phase 2</u></p> <p>Case Study Reports</p> <ul style="list-style-type: none"> • Key stakeholder interviews and country visits • Case studies in six countries (5 pages each) 	20%
Deliverable 3	April 2021	<p><u>Phase 3</u></p> <p>Draft Evaluation Report, including the following content in draft form:</p> <ul style="list-style-type: none"> • Executive summary • Literature review (from inception phase) • Final methodology including sampling strategy • Analysis of key findings for each evaluation question • Conclusions and Lessons Learned • Recommendations, country specific and regional • Bibliography (from inception phase) 	40%

		Validation Workshop of preliminary findings, conclusions and prioritization of recommendations from draft report.	
Deliverable 4	May-June 2021	Final Evaluation Report, Brief, Video and Presentation , including the following content: <ul style="list-style-type: none"> • Executive summary • Literature review (from inception phase) • Final methodology • Analysis of key findings for each evaluation question • Conclusions and Lessons Learned • Recommendations, country specific and regional • Annotated bibliography (from inception phase) • A 4-page brief intended for a broader, non-technical audiences • A PowerPoint presentation and video for UNICEF and SEAMEO to use for general audiences 	20%

*Deadlines and timelines are subject to change.

6.6 Evaluation dissemination and utilization plan

As stated above, the evaluation team will conduct a validation workshop of preliminary findings, which will serve to both valid the findings and open up a discussion on the evaluation findings with key stakeholders. In addition, the evaluation team will provide a final evaluation report with a 4-page brief and a PowerPoint presentation and video for UNICEF and SEAMEO to use for general audiences. Beyond the delivery of these deliverables, resources have not been assigned to the dissemination of the evaluation findings for the evaluation team. Therefore, it is foreseen that any evaluation dissemination and utilization plans will be led and implemented by UNICEF and SEAMEO.

7 Risk management and risk mitigation

The evaluation team will make sure that risks are mitigated through upstream planning and pre-emptive mitigation to maximise the likelihood that the evaluation will achieve its objectives. It will be the responsibility of the Project Manager to review and update the risk matrix presented below ('living documents') and the corresponding mitigation strategies on a regular basis. The Project Manager will keep UNICEF ESARO aware of changing risk levels and propose and implement mitigation and management strategies.

Risk is unavoidable, and the resources to manage risk are also finite, so it is important to constantly evaluate all risks and keep UNICEF informed. It will be imperative for us to manage and mitigate those risks that are within our sphere of control. Table 4 outlines risks, with probability, impact and mitigation options. These were presented at the proposal stage and remain risks that will continue to be monitored.

Table 4 Risks, probability, impact and mitigation

Risk	Probability	Impact	Mitigation
CE project staff do not deliver agreed outputs (on time, on budget, as agreed in ToR)	L	H	Quality assurance system established and applied in all project areas. Ongoing liaison with UNICEF and quality assurance of all deliverables.
Programme activity behind schedule	M-H	H	Activity workplan agreed upon and approved by UNICEF; clear communication between CE's Project Manager and UNICEF main contact to discuss any delays or issues identified; effective communication to ensure that deadlines and progress are visible to all teams.
Tight timelines require CE to proceed with work without UNICEF approval. E.g. Synthesis of case studies to begin prior to feedback on country reports provided.	H	M	CE to be transparent about work that has begun prior to UNICEF approval of documents; UNICEF to endeavour to provide feedback as rapidly as possible; CE to remain flexible about amendments/changes that might occur. Turnaround time to be agreed for comments, feedback and approval of documents.
CE consultants do not perform as agreed	L	H	All consultants will be carefully chosen based on their experience in delivering similar assignments. CE will closely monitor performance and rearrange the team if necessary, in consultation with UNICEF.
Risk of scope creep and overstretch on project team	L	M	Implement activities with clear, agreed deliverables and timeline broken down to achievable targets; ensure constant dialogue with UNICEF colleagues.
Difficulty in identifying and accessing the appropriate people and	L	L	CE will work closely with UNICEF's EAPRO and through the office, with UNICEF's relevant country offices. Cambridge

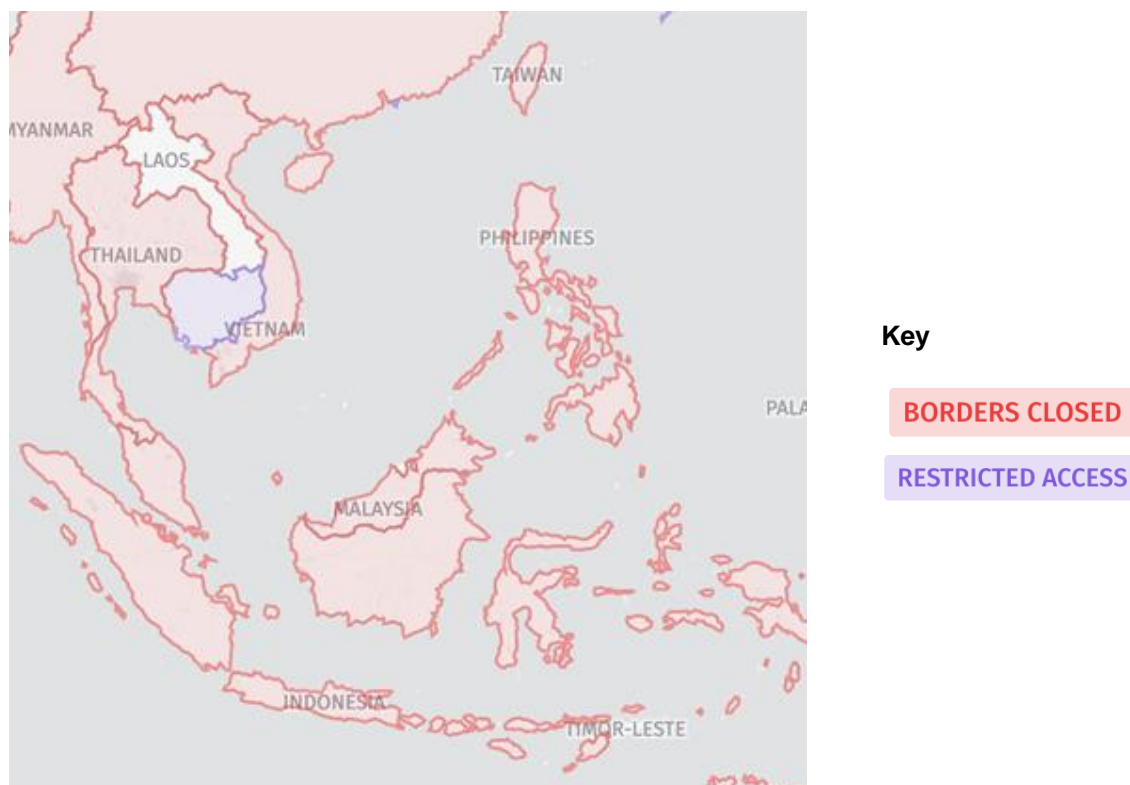
Risk	Probability	Impact	Mitigation
documents, due to Cambridge Education not having offices in countries falling within the evaluation			Education will use UNICEF's logistical support and rely on UNICEF's local knowledge and contacts to source documentation, identify appropriate interviewees, arrange meetings and ensure no negative impact of our lack of permanent presence in the countries where the evaluation is carried out.

7.1 Realities and scenarios - COVID-19

Since the proposal writing stage, the global pandemic COVID-19 has affected almost every country. The pandemic brings significant risks to the project, some of which can be mitigated and others which cannot. During the global pandemic, the inability of evaluation team members to travel to the UNICEF Regional Office and the case study countries is now, not only a risk but a certainty. Thailand and Malaysia have closed their borders and Cambodia has restricted access to visitors. In addition, the Foreign and Commonwealth Office in the UK (where the Team Leader resides) has advised against all non-essential travel. Therefore, undertaking face-to-face inception workshops and missions within the country case studies immediately after inception is not possible.

It was agreed with UNICEF to continue through the inception period and to conduct a virtual inception workshop and that the inception report would provide a set of scenarios for the evaluation, in the time of COVID.

Figure 63 Border closures in Asia



Scenario A: One option would be to postpone primary data collection to a time where travel is possible. The obvious advantage to this approach is that nuances are often more easily picked up through missions and face-to-face communication. However, there are several disadvantages to Scenario A also. Firstly, depending on the workplans of participating countries and regional stakeholders, in light of COVID-19, the findings from the evaluation may not be available to stakeholders in a timely manner to fully inform the planning for the next phase of SEA-PLM. As this is the key objective of the evaluation, this is a large risk. Secondly, respondents may find it more challenging to accurately remember and reflect on the activities under review in this evaluation the further that these activities move into the past. And thirdly, when nation's open up and travel is possible, the attention of key stakeholders may well be diverted to other priorities. If the data collection was to be delayed to such an extent that the results from the PLM surveys became available during the data collection and analysis period, it may mean that the scope of the evaluation could be extended to include a more thorough look at impact. To enable this to happen, the evaluation matrix would need to be extended and significantly more data analysis would have to be conducted. This would require considerably more resources than are currently available.

Scenario B: If travel and movement restrictions continue, our preference would be moving ahead with the evaluation and relying on remote data collection techniques. As at the time of finalising this Inception Report, Jeff Marshalls is resident in Laos (one of the case study countries). Laos has recently reduced some of its lock-down measures and although donor partners are mainly still working from home, government officials have begun to resume work in their offices. It is therefore possible (putting safety considerations first) that some face-to-face meetings could be conducted for this case study (bearing in mind safe distancing and other health guidelines). It is also important to note that one of the reasons for selecting Jeff for this piece of work, is his long experience working on assessment in the sub-region. Jeff has over ten years of recent experience working in Cambodia, which means that remote calls with key stakeholders will be less of a challenge for this case study country as some of the disadvantages mentioned below (such as judging response or analysing the political economy) will be minimised due to the relationships Jeff already has with key informants.

The obvious advantage for Scenario B is the ability to move forward with the evaluation and provide findings within a timeframe most likely to be able to support decisions regarding the next phase of SEA-PLM. The team has proposed several mechanisms to overcome the identified disadvantages (see below). The evaluation team has added a regional online survey in addition to the initially agreed scope for the evaluation, in recognition of the need to elicit as much information as possible, remotely.

In Table 5 below, we have presented two potential scenarios where movement restrictions are in place and how we could potentially respond.

Table 5 Implications of COVID: Scenarios A and B

	Scenario A: Delay primary data collection until travel is possible	Scenario B: Virtual missions
Desk Research and Inception report	Timing as planned	Timing as planned
Inception workshop	Timing as planned. Virtual workshop.	Timing as planned. Virtual workshop.

	Scenario A: Delay primary data collection until travel is possible	Scenario B: Virtual missions
Primary qualitative data collection	Indeterminately delayed.	Timing as planned. Interviews conducted virtually.
Advantages	Face to face interviews support the capture of context and relationships more easily.	Work can be completed within the planned timeline and inform the next stage of SEA-PLM planning. Accessing respondent may be easier as many people are now working from home.
Disadvantages	Accessing respondents may be more difficult once borders open and schools ease restrictions. The evaluation may not be available to inform the next stage of SEA-PLM programming at a crucial time.	Internet connections may be a challenge in some cases, resulting in frustrating calls with respondents unable to communicate their message clearly. Purposive actions to mitigate loss of nuanced contextual information would need to be introduced. Scheduling and conducting interviews is likely to take more time.

7.1.1 Mitigating the challenges of remote research

Below we consider potential challenges with conducting online/phone interviews and how we propose to adapt to mitigate the challenges of remote research:

1. Accessing respondents may be more difficult.

After being introduced to proposed respondents, we will contact them well in advance of the proposed interview being conducted. We will follow up with respondents through a structured and recorded (in an excel document) weekly follow up email and when possible, reach out via phone to schedule interviews.

2. It is more difficult to judge body language, tone or other non-verbal signals, more difficult to build rapport and respondents may tend to give shorter answers over the phone or be less candid.

We will begin each interview with an introduction to the evaluation and a brief outline of the purpose of the interview. Phone interviews and conversations can be frustrating when expectations for the purpose of the conversation are not aligned. Next, the evaluator will begin a discussion on the respondent's current role and experience in the education sector and seek to build rapport early. The evaluator will reflect the words and sentiments of the respondent back after several sentences in order to demonstrate that the evaluator is listening. The evaluator will match the speed at which the respondent speaks as this often reflects the speed at which they are comfortable engaging. Where the evaluator senses the respondent may take longer to 'warm' to the situation, a follow-up interview will be suggested so more time can be spent on rapport without the respondent experiencing fatigue.

3. The interviewing schedule is likely to be less efficient. There are likely to be some disruptions or connectivity issues, and time zone differences need to be taken into account.

Rather than conducting each (online) country mission within a 5 day period, after the initial pilot, the evaluation team will conduct the interviews over a longer calendar period to allow for follow up interviews and in order to ensure the right people and a sufficient number of people are interviewed.

4. Face to face missions more easily capture informal relationships and aspects of the political economy which may influence the programme.

These aspects can be captured virtually, but it does take more effort. As experienced evaluators, we have faced similar circumstances in the past and during COVID on other evaluations. We have developed methodologies that can help to mitigate some of the constraints of remote/virtual data collection.

For example: We would introduce researcher reflection and debriefs after each interview. This often takes place organically during a mission rather than in a routine and structured way. We would develop a set of notes for each country and ensure that the notes are reviewed before each interview and added to after each interview with stakeholders in the specific country. In this way the evaluator will be immersed in the context prior to each interview and will capture any nuances during the interview.

5. Remote research and online video calls have been around for years and there are many different tools to choose from. Problems can arise when every platform seems to need installing and a new account created. This can be time consuming and can be frustrating for respondents.

We will use zoom and ensure respondents can join the call online without having to set up an account. We will provide a step-by-step guide to respondents on how to join the call in advance of the call.

6. Respondent communication can be misunderstood or lack nuance where there are connectivity issues.

The evaluation team has added an online survey to the data collection activities to ensure key pieces of information are captured. In addition, after each interview the evaluator will send notes of the interview to the respondent over email and provide an opportunities for the respondent to correct, clarify or further nuance any of the communication.

7.2 Additional risks introduced by COVID-19

In addition to introducing the need for scenarios for conducting the evaluation, two major additional risks are introduced by COVID-19, under both scenarios. Below we outline these risks and how we plan to mitigate them.

Risk 1: The COVID-19 situation is quite different in each of the four countries and the extent to which regional and national government officials will be both contactable and able to prioritise the study in the times of COVID brings risks.

As at 12 May 2020, Thailand's lockdown is beginning to ease, with restaurants, markets, parks and sporting venues allowed to reopen, but with a curfew of 10pm in place. Malaysia is in lockdown at least until June 9, although cafes and restaurants can reopen but must practice social distancing. School are closed and borders are shut. In Laos, the freedom of movement for people residing in Laos is increasing, with restaurants and cafes, and night markets reopened. Schools are reopening gradually for some grades. All borders are closed. Lockdowns

and school closures have been put in place in Cambodia and there is restricted access for visitors and bans for visitors from several countries are in place.

To mitigate this risk, we will contact key informants for the evaluation as early as possible in the project implementation period and re-contact weekly, to provide the highest chance of reaching them and scheduling an interview.

Risk 2: The risk that plans for SEA-PLM beyond the main survey data collection will be disrupted by COVID is a major risk. This could disrupt the utility of the recommendations proposed in the final report.

This risk cannot be fully mitigated and reflects the levels of uncertainty in all activities in the present moment. In order to mitigate the risk as much as possible, we propose to include future scenarios in both the collection of information and in the presentation of recommendations to UNICEF. KII data collection tools will include questions on the likely future scenarios for SEA-PLM activities in each country and the analysis of primary and secondary data will cast an eye to the potential future opportunities and threats.

8 Bibliography

- ACER, SEAMEO & UNICEF. (2017). *Southeast Asia Primary Learning Metrics (SEA-PLM) Assessment Framework*. Melbourne: ACER.
- Ahmad, A. e. (2015). A critical analysis of the evolution of the Southeast Asia Primary Learning Metric (SEA-PLM). *Learning for sustainable futures: Making the connections*. Oxford: UKFIET.
- 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*. London: EEPI-Centre, University of London.
- Baird, Jo-Anne, et al., *Policy Effects of PISA*, Oxford University Centre for Educational Assessment, Oxford, May 2011
- Bernard, Jean-Marc., *Managing the Impact of PASEC Projects in Francophone Sub-Saharan Africa*, CONFEMEN secretariat, Hamburg, 2006
- Breakspear, Simon., *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance*, OECD, Paris, Feb 2012
- Breakspear, Simon., *How does PISA shape education policy making? Why how we measure learning determines what counts in education*, Centre for Strategic Education, Melbourne, Nov 2014
- DFAT. (2020a). *DFAT Cambodia*. Retrieved from DFAT : <https://www.dfat.gov.au/geo/cambodia/Pages/cambodia-country-brief>
- DFAT. (2020b). *Overview of Australia's aid program to Myanmar*. Retrieved from DFAT: <https://www.dfat.gov.au/geo/myanmar/development-assistance/Pages/development-assistance-in-myanmar>
- Global Partnership for Education . (2020e). *Where we work. Myanmar*. Retrieved from Global Partnership for Education: <https://www.globalpartnership.org/where-we-work/myanmar>
- Global Partnership for Education. (2020a). *Where we work. Cambodia*. Retrieved May 2020, from Global Partnership for Education: <https://www.globalpartnership.org/where-we-work/cambodia>
- Global Partnership for Education. (2020b). *GPE's expanding role to improve education in the Asia Pacific*. Retrieved from Global Partnership for Education: <https://www.globalpartnership.org/blog/gpes-expanding-role-improve-education-asia-pacific>
- Global Partnership for Education. (2020c). *Where we work. Lao PDR*. Retrieved from Global Partnership for Education : <https://www.globalpartnership.org/where-we-work/lao-pdr>
- Global Partnership for Education. (2020d). *Where we work. Vietnam*. Retrieved from Global Partnership for Education: <https://www.globalpartnership.org/where-we-work/vietnam>
- HEART., *The impact of assessment results on education policy and practice in East Africa*, DFID think piece, London, 2016

- Howie, Sarah. and Plomp, Tjeerd., *International comparative studies of education and large-scale change*, chapter in *International Handbook of Educational Policy*, 2005, p75-99.
- Iyer, P., Azubuike, O. B., & Rolleston, C. (2017). *Young Lives School Survey 2016-2017*. Oxford: Young Lives.
- Lietz, P. & Tobin, M. *The impact of large-scale assessments in education on education policy: evidence from around the world*, Research Papers in Education, 2016
- Loizillon, A., & Uys, P. (2020). *EAPRO Meta-analysis on what works to improve learning outcomes in primary education in East Asia and the Pacific. Literature Review: Learning Outcomes in East Asia and the Pacific (Volume 1)*.
- Marivin, A. (2019). *Matrix analysing funding needs and priorities in UNICDEF and COs and roadmap to support COs, particularly GPE countries*.
- Monteiro, A. R. (2010). The right of the child to education: what right to what education? *Procedia Social and Behavioral Sciences*, 1988-1992.
- Mullis, I. V., Martin, M. O., Foy, P., & Hooper, M. (2016). Retrieved from TIMSS 2015 International Results in Mathematics. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2015/international-results/>
- Murimba, Saul,. *Evaluating students' achievements: the impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)*, Prospects, 2005, p91-108.
- n.p. (2015, July 16). South East Asia Primary Learning Metrics background, rationale and programme logic. *South East Asia Primary Learning Metrics background, rationale and programme logic*. n.p.
- OECD. (2020). *OECD-DAC*. Retrieved from Receipts: https://public.tableau.com/views/OECDDACaidataglacebyrecipient_new/Recipients?:embed=y&:display_count=yes&:showTabs=y&:toolbar=no?&:showVizHome=no
- OXFAM. (2019). *Financing for Development in Vietnam*. Oxford: Oxfam.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
- SEAMEO and UNICEF. (2019b, October 10). *Meet the new Technical Advisory Group for SEA-PLM*. Retrieved from SEA-PLM website: https://www.seaplum.org/index.php?option=com_k2&view=item&id=15:meet-the-new-technical-advisory-group-for-sea-plm
- SEA-PLM. (2015). *Southeast Asian Primary learning Metrics. Background, Rationale and Programme Logic*. SEA-PLM.
- SEA-PLM. (2019a). *SEA-PLM 2019 Technical report*.
- SEA-PLM. (2019b). *Target Audience Matrix*.
- SEA-PLM SEAMEO UNICEF . (2019a). *Regional and national activity plan for supporting analysis, reporting, dissemination and use of results of SEA-PLM*.

- SEA-PLM SEAMEO UNICEF. (2019b). *Communication plan for the SEA-PLM Secretariat to disseminate and support the use of results, methodology and data at regional and national level for SEA-PLM 2019*.
- Spink, J. (2019). Policy Priorities and Possibilities for Data Analysis. SEA-PLM.
- Stanley, G.,. *Reforming educational assessment: Imperatives, principles, and challenges*, Australian council for Education Research (ACER), Melbourne, 2012
- The ASEAN Secretariat. (2016). *Initiative for ASEAN Integration (IAI) Work Plan III*. Jakarta.
- UNESCO Asia and Pacific Regional Bureau for Education. (2018, June 15). *Better and more information on factors influencing learning across Southeast Asia*. Retrieved from UNESCO Bangkok: <https://bangkok.unesco.org/content/better-and-more-information-factors-influencing-learning-across-southeast-asia>
- UNICEF. (2007). *A Human Rights-Based Approach to Education for All*. New York: UNICEF.
- UNICEF. (2019). *Final SEA-PLM Hand-over notes*.
- UNICEF. (2019). *Terms of Reference for an Evaluation of the SEA-PLM Programme in Southeast Asia*. Bangkok: UNICEF Bangkok.
- UNICEF and SEAMEO. (2019a). *SEA-PLM 2019 Assessment Framework (1st ed.)*. Bangkok, Thailand: United Nations Children's Fund (UNICEF) & Southeast Asian Ministers of Education Organization (SEAMEO) - SEA-PLM Secretariat.
- UNICEF EAPRO. (2020, May 6). *Education: Every child has the right to go to school and learn*. Retrieved from UNICEF ESAPRO Education Programme: <https://www.unicef.org/eap/what-we-do/education>
- UNICEF EAPRO. (n.d.). *Terms of Reference for SEA-PLM Advisor*. Bangkok: UNICEF EAPRO.
- UNICEF EAPRO. (n.d.). *TOR for SEAPLM Team - UNICEF and SEAMEO Secretariat*.

Annexes

Annex A - Terms of Reference

TERMS OF REFERENCE FOR INSTITUTIONAL CONTRACTS.



UNICEF Bangkok

Requesting Section: EDUCATION/EVALUTION SECTION, EAPRO

TITLE: Terms of Reference for an Evaluation of the Southeast Asia Primary Learning Metrics (SEA-PLM) Programme in Southeast Asia

1- Background and rationale

The Regional Context in Southeast Asia

UNICEF East Asia and the Pacific Regional Office (EAPRO) is supporting a regional large-scale learning assessment with the Southeast Asian Ministers of Education Organization (SEAMEO) since 2012. The Southeast Asia Primary Learning Metrics (SEA-PLM) is the first regional learning assessment at primary level (Grade 5) in Southeast Asia – www.seaplm.org.

As an alternative to international large-scale assessments initiatives (OCDE-PISA, IEA-PIRLS, TIMSS) and national tools (EGRA, national assessments) already used in most of the Southeast Asian countries as part of their system-level mechanisms to generate robust learning data on national representative sample thanks to standardized testing. SEA-PLM is designed as a regional Programme of cyclic assessments to measure quality and equity trends in education over time, sub-population and countries by repeating cycles. This regional initiative is a step forward for measuring, accelerating and tracking progress over SDG 4.1.1 (b) and 4.7.4 by defining and reporting collected data on common scales, levels and learning descriptors.

SEA-PLM collects learning data on national representative samples of students as required by the Global Partnership in Education and the Unesco Institute for Statistics for reporting learning progress. SEA-PLM is integrated into international mapping as one of the active Regional Large-Scale Assessments (ERCE, PASEC, SEA-PLM, SACMEQ, PILNA). Regional assessments are considered by the international community as critical tools to track SDG 4.1 targets over time thanks to their nationally representative samples and reliable competencies scales.

The implementation of each cycle of the assessment is framed to follow a generic framework and timeline of 4-5 years standardized with participating countries. The SEA-PLM first round is branded SEA-PLM 2019 in reference to the data collection year and is having an average duration

period of 5-6 years (2015-2020/2021 for SEA-PLM 2019). Six⁹ (6) of the eleven¹⁰ (11) Southeast Asia countries have decided to implement SEA-PLM 2019 survey, however, all Southeast Asia countries are involved in the Programme governance structure, methodology development and policy exchange. Those countries may join the next round of the assessment. The initial methodological framework and procedures and SEA-PLM Programme inception have been developed between 2012 and 2015 with the contribution of all countries.

GPE countries use results from Regional Large-Scale Assessment to report national progress in learning outcomes, actors' practices, expectation and attitudes, and evaluate project impact on targeted groups within specific domains and context. Such assessments are used as a strong component of Education Sector Plans to monitor, plan, design and refine policies and strategy documentation related to curriculum, assessment, resources allocation, equity, workforce development and accountability.

SEA-PLM focuses also on supporting Southeast Asia countries by developing a network and community of practices, mentorship and peer-learning between experts and countries to better measure and understand the status of student learning achievement and thereby improve the quality of their education systems. The regional collaborative framework offers the possibility to build individual and collective capacities around learning policies and assessment through SEAMEO and ASEAN agenda.

SEA-PLM expects to achieve three key outcomes including:

1. Enhanced capacity to generate and analyse assessment data at regional, national and sub-national levels.
2. Enhanced capacity to utilize assessment data for education improvement and more equitable learning outcomes at regional, national and sub-national levels
3. Enhanced ASEAN integration in terms of approaches to assessment, with an initial focus on primary Grade 5 in the domains of numeracy (mathematics), literacy (reading & writing), and global citizenship.

The initiative gathers multiple actors and stakeholders from the regional and national level on time-consuming timelines, standardized in-countries survey activities, high-level expertise for methodology development and quantitative/qualitative research, political commitment and transformation of results in educational policies and practices. The Programme is also engaged in international and regional learning policy dialogue on SDGs, learning policies and large-scale assessment area.

The direct beneficiaries of the assessments and policy exchanges are the Ministry of Education officials, the Education policymakers and sector planners. The development partners, researchers in education and technical institutions get also the opportunity to take benefit from quality data, capacity building in assessment and improved standards across the region. At bottom level, all current and future children studying at primary level and teachers in the region will benefit from the future education reforms that are informed by SEA-PLM findings for improving learning opportunities and professional practices.

⁹Cambodia, Malaysia, Myanmar, Lao PDR, Philippines and Vietnam.

¹⁰ Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste and Vietnam.

ASEAN, SEAMEO and other regional and international organizations will also benefit from increased coordination, alignment and increasing Southeast integration at regional and international level as Global Alliance for Monitoring Learning (GAML).

UNICEF EAPRO and UNICEF CO are supporting SEAMEO Secretariat and participating ministries in their Programme participation by supporting the institutional development of the Programme, mobilizing external expertise to undertake and control technical operations and contribute to funding national and regional operation for training, data collection, communication and human resources. The day to day management, reporting and decision making is jointly managed by UNICEF EAPRO and SEAMEO Secretariat through different bodies and actors at the regional and national level.

Time to take stock: Context for the external evaluation

With the coming implementation of the last stages of the first round of the first regional large-scale assessment in Southeast Asian Region, the Programme is going to deliver SEA-PLM 2019 by reporting, disseminating and supporting the use of results in 2020 and 2021. In the meantime, the Programme is continuing to transform this approach toward a long-term Programme.

The necessity of exploring strength and weakness of the first assessment on a governance, technical, managerial, funding and partnership perspectives, is necessary before engaging in the second round of the assessment and switching from a biannual plan and commitment to a 5 years approach to support a complete round of assessment and increase the stability, visibility and ownership of the project.

As one of the Co-Chair of the Secretariat of SEA-PLM and main donors of the assessment, UNICEF wants to commission an independent evaluation of this intervention.

The evaluation will focus both on (i) the inception phase of the assessment (2012-2015) and (ii) the implementation of the first round of the assessment (SEA-PLM 2019). The evaluation primary focus will be on SEA-PLM 2015 to 2019. SEA-PLM has the potential to organize three rounds of assessment before 2030 SDG's.

2- Purpose, Objectives & Scope

The main purpose of this independent evaluation is to draw lessons learned and account for results to inform the next phase of the SEA-PLM Programme. As such it combines a formative and summative focus and intends to inform future Programme planning and re-positioning. To do so, this evaluation will assess the relevance, effectiveness, efficiency, impact and sustainability as well as equity, gender equality and human rights considerations of the SEA-PLM Programme. It will also consider these criteria in relation to issues including national coordination, leadership and institutionalization according to national priorities and global frameworks. The evaluation will draw evidence-based findings and recommendations which will identify both emerging good practices and new regional strategic approaches, programmatic components and intervention modalities needed to better position UNICEF and its partners in a fast-changing EAP regional context in support of national governments in achieving their national goals and targets and contribute to the SDG agenda.

The primary audience of this evaluation is UNICEF EAPRO, SEAMEO, the SEA-PLM Secretariat, Country Office Management teams and Education teams; the secondary audience includes governments' development partner, donors and external stakeholders as contractors and other collaborators. The findings, conclusions, and recommendations will be used by both the regional and country offices to inform the development or implementation of their country Programme documents and adjust their strategies and implementation modalities as necessary.

At the regional level, it will inform the design and the implementation of a SEA-PLM long-term strategy and multi-annual activity plan as the regional learning Strategy of the UNICEF regional office, and as well as the regional policy advocacy and partnerships leveraging efforts. Findings of the evaluation will also be disseminated to other UNICEF Regional Offices and UNICEF Headquarters.

The objectives of this evaluation are to:

1. Review the relevance, effectiveness, efficiency, impact, the sustainability of the SEA-PLM Programme and assess the extent to which it has been implemented in line with the objectives and global good practices on regional learning assessments;
2. At regional and national level, identify what strategic approaches, programmatic components and intervention modalities can help better position UNICEF and collaborators for SEA-PLM Programme and much broader in a fast-changing EAP region and global initiatives and in support of country programming in achieving the SDG agenda.

This evaluation will also strengthen and update the knowledge and evidence base on SEA-PLM in complementarity with other initiatives recently completed, on-going or recently launched by UNICEF in the region such as the literature review of what works to improve learning in the region.

Evaluation Scope

- The evaluation will be retrospective (*summative*) and cover (*secondary*) the conception period of SEA-PLM (2012-2015) and (*primarily*) the implementation of SEA-PLM 2019 from 2015 to 2019, but it will be forward-looking (*formative*) in providing conclusions and recommendations for regional- and country-level programming.
- Time boundaries: from 2012-2019 with a focus of the current implementation period of SEA-PLM 2019 (2015-2019).
- Organizational level: UNICEF country offices covers by SEA-PLM, UNICEF regional office and the Regional Secretariat and a focus of SEA-PLM 2019 participating countries and consultation of non-participating countries.
- The geographic scope of the assignment will cover six (6) Southeast Asian countries participating in SEA-PLM 2019: Cambodia, Lao PDR, Malaysia, Myanmar, Philippines, and Vietnam. The desk review will cover eleven (11) SEA countries.
- Type of interventions: both upstream (enabling environment and system strengthening work as learning and/or assessment policies planning, implementing and monitoring) and downstream programming strategies (enabling local environment and actors as learners, teachers, head teachers and community in developing capacities and new opportunities for teaching and learning).

Evaluation framework and questions

Evaluation evidence will be drawn against the Organisation for Economic Co-operation and Development, Development Assistance Committee's criteria of relevance, effectiveness, efficiency, sustainability and impact. These criteria are prioritized because they capture the evaluation questions presented below. In addition, the evaluation will review equity and gender as cross-cutting issues.

Key evaluation questions (and sub-questions) are clustered according to the evaluation criteria provided. This initial list of questions will be further refined and unfolded by the evaluation team and included in the Inception Report following desk review of key documents.

Relevance of SEA-PLM to the priorities and policies of national education stakeholders in the eleven (11) Southeast Asian countries

1. To what extent is SEA-PLM aligned to i) the national development policies and priorities of national education stakeholders and ii) UNICEF's national, regional and global objectives and intended impacts?
2. What are the programming gaps or unaddressed needs? What could be done better?

Effectiveness of SEA-PLM to intended outcomes in the six (6) SEA-PLM 2019 participating countries

3. To what extent have the objectives and expected outcomes of SEA-PLM been achieved or are likely to be achieved?
4. What have been the major factors influencing the achievement or non-achievement of SEA-PLM objectives and activities? What were the enabling factors, barriers and bottlenecks?
5. What can UNICEF and its partners do to ensure the objectives of SEA-PLM are met in the future? What kind of initiatives should UNICEF prioritize at the country and regional?

Efficiency of the management of SEA-PLM and timely use of resources at country and regional levels in the six (6) SEA-PLM 2019 participating countries

6. How well have SEA-PLM activities been managed by UNICEF in terms of the technical and financial resources? Have SEA-PLM activities been implemented in the most cost-efficient way compared to alternative approaches?
7. What were the strengths and weaknesses of SEA-PLM management processes? How could management of SEA-PLM activities be improved? To what extent and how did the UNICEF Regional Office contribute to efficiency of SEA-PLM in education programming?

Impact resulting from the SEA-PLM Programme at this stage of its implementation (positive and negative changes, intended and unintended) in the six (6) SEA-PLM 2019 participating countries

8. How has SEA-PLM contributed to national education systems and assessment practices and policies and discourses thus far? And at the regional level, what positive and negative changes has SEA-PLM brought about?

Sustainability of SEA-PLM and benefits

9. To what extent can SEA-PLM activities, plans and strategies be fully integrated and implemented by the government (s), both technically and financially? To what extent are they likely to continue?
10. What are the key barriers and bottlenecks towards achieving sustainability of SEA-PLM activities?
11. How can SEA-PLM attract other countries in Southeast Asia and better link other international and regional initiatives?

Equity and gender equality considerations

12. To what extent is SEA-PLM conducive to supporting the most marginalized populations and genders (including those furthest left-behind)?

These questions can be further refined by the evaluation team and included in the Inception Report following desk review of key documents.

3- Evaluation approach and methodology

The primary task of the consultant is to undertake an evaluation of the SEA-PLM Programme to date and provide formative guidance to inform the future developments of the Programme.

It is expected that the evaluation will employ a non-experimental design and use both a theory-based (re-constructing the theory of change), utilization-focused, participatory and a mixed-methods approach drawing on key background documents, monitoring frameworks at country and regional levels and interviews with key informants. Key documents, data and a contact list of relevant informants will be provided to the evaluation team once a contractual agreement has been made. Documentation will include methodological materials, ToR and meeting minutes of bodies of governance, activity plans, letter of commitment and political statement as an example.

The methodology will primarily use mixed data collection methods include:

- **Brief Literature review** of emerging good practice concerning regional and global large-scale learning assessments;
- **Desk review** of key SEA-PLM Programme documents;
- **Key informant interviews** with SEA-PLM technical team, SEAMEO, UNICEF staff at regional and global levels, and key counterparts at national, regional and global level considering the eleven (11) SEA countries; and,
- **Six (6) Case studies** covering SEA-PLM 2019 participating countries

A limitation to the evaluation is the timeframe. Because the results of SEA-PLM 2019 will only be released in 2020, the evaluation will not be able to assess impact of results on children and education systems. Rather it will take a more formative approach, looking at intermediate outputs and outcomes. The applicants should discuss the above or other potential limitations in their proposal.

Based on the above-mentioned purpose and scope of the evaluation, the bidder is expected to submit a proposal for methodologies including data sources and selection criteria among regional and country programming strategies as well as the set of key evaluation questions. Methodological rigor will be given significant consideration in the assessment of proposals. Hence bidders are invited to interrogate the approach and methodology preferred in the ToR and improve on it, or propose an approach they deem more appropriate. In their proposal, bidder should clearly refer to triangulation, sampling plan and methodological limitations and mitigation measures. Bidders are encouraged to also demonstrate methodological expertise in large-scale assessment.

Country visits are envisaged to cover six countries case studies.

The evaluation will occur in three phases, which are further described below: 1) Inception phase, 2) Data collection and analysis phase, and 3) Reporting and communication phase. The following offers guidance on the evaluation process and UNICEF's expectations and thinking. It should be commented on, further developed and improved by the bidders in their respective proposals. Alternative approaches can also be proposed. The methodology will be further specified and finalized by the selected evaluation team in collaboration with UNICEF during the inception phase.

Phase 1. Inception Phase

Step 1. Literature search and review. The evaluation team will review relevant data and analyses from national, regional and global frameworks to understand the context in SEA-PLM were and are operating, and research and practical evidence on successful approaches to regional and global large-scale learning assessment relating to national framework to monitor learning. The literature review should form the basis of the background and context section of the evaluation.

Step 2. Desk review. The evaluation team will review key SEA-PLM background documents from the eleven countries and the regional level, including assessment framework, technical standards, Secretariat strategy documents, monitoring reports, communication materials and field trial documents. The aim of the desk review is to familiarize the consultants with SEA-PLM, to contribute to the background and context sections of the report and to shape the methodology for the inception report.

Step 3. Preparation of Inception Report that includes evaluation methodology and tools. The methodology should be prepared to cover all the intended objectives of the evaluation, including: Summary of literature and desk review, reconstructed theory of change, evaluation criteria and questions, methodology and data collection and analysis tools, foreseen limitations, list of priority interviewees, workplan with timeline and bibliography. The evaluation methodology design and evaluation questions will be finalized in agreement with the UNICEF Regional Office.

Phase 2. Data collection and analysis phase

Step 4. Data collection. The evaluation team will collect data using a mixed methods approach. Data will be collected remotely from all the six countries. Field visits will be included in bidders' proposals for covering the six countries.

Step 5. Data analysis. The evaluation team will compile and analyze data. The methodology used should be clearly described in the final report.

Step 6. Case studies and preliminary findings. The evaluation team should prepare a case study of about 5 pages for each country participating in SEA-PLM to document the evolution of SEA-PLM in the country and its current role and contributions to the education sector as well as gaps, opportunities and recommendations for maximizing impact and long term Programme sustainability.

Phase 3. Reporting and communication phase

Step 7. Draft report. The consultant will prepare a draft report, with conclusions, lessons learned and recommendations for the way forward drawn from the analysis. The report structure should follow [the UNICEF-Adapted UNEG Evaluation Report Standards \(2017\)](#).

Step 8. Validation workshop with UNICEF, evaluation reference group, and other stakeholders (validation workshop if applicable, virtual meetings) to validate findings, conclusions and prioritize recommendations.

Step 9. Finalisation of the evaluation report. The consultant will present the final draft evaluation conclusions and recommendations to SEAMEO and UNICEF RO, using a PowerPoint presentation and/or other methodologies. Comments and feedback on the findings and recommendations should be incorporated to finalise the report.

The evaluation methodology should follow the [UNICEF Evaluation Policy \(2018\)](#), [the United Nations Evaluation Group \(UNEG\) Norms and Standards for Evaluation \(2016\)](#), [UN SWAP Evaluation Performance Indicator](#), [UNEG Ethical Guidelines \(2008\)](#), [UNICEF Procedure for Ethical Standards and Research, Evaluation and Data Collection and Analysis \(2015\)](#) (Report must be compliant with [the UNICEF-Adapted UNEG Evaluation Report Standards \(2017\)](#))

4- Deliverables and payment schedule

No.	Tentative Deadline	Deliverable	Percent payment
Deliverable 1	March 2020	<p><u>Phase 1</u></p> <p>Inception Report and presentation, including:</p> <ul style="list-style-type: none"> • Literature review • Desk Review • Methodology and data collection and analysis tools including sampling strategy • Foreseen limitation and mitigation measures • Workplan with timeline • Annotated Bibliography • Presentation to SEAMEO and UNICEF EAPRO (remotely) 	10%

Deliverable 2	May-June 2020	<p><u>Phase 2</u></p> <p>Case Study Reports</p> <ul style="list-style-type: none"> • Key stakeholder interviews and country visits • Case studies in six countries (5 pages each) 	10%
Deliverable 3	August 2020	<p><u>Phase 3</u></p> <p>Draft Evaluation Report, including the following content in draft form:</p> <ul style="list-style-type: none"> • Executive summary • Literature review (from inception phase) • Synthesis of desk review (from inception phase) • Final methodology including sampling strategy • Analysis of key findings for each evaluation question • Conclusions and Lessons Learned • Recommendations, country specific and regional • Bibliography (from inception phase) <p>Validation Workshop of preliminary findings, conclusions and prioritization of recommendations from draft report.</p>	30%
Deliverable 4	November 2020	<p>Final Evaluation Report, Brief, Video and Presentation, including the following content:</p> <ul style="list-style-type: none"> • Executive summary • Literature review (from inception phase) • Synthesis of desk review and mapping of country level activities (from inception phase) • Final methodology • Analysis of key findings for each evaluation question • Conclusions and Lessons Learned • Recommendations, country specific and regional • Annotated bibliography (from inception phase) • A 4-page brief intended for a broader, non-technical audiences • A PPT and video for UNICEF and SEAMEO to use for general audiences 	50%

*Deadlines and timelines are subject to change.

5- Reporting requirements

The deliverables listed above should include the reporting requirements, as described in the table above. The final report should follow UNICEF guidelines and be no more than 40 pages (or 12,000 words excluding annexes) and should include an executive summary. The final report should also include a summary PPT presentation.

In addition, regular progress reports should be provided either via email or phone to the UNICEF RO team, as requested.

All deliverables must be in professional level standard English and in compliance with [UNICEF Style Book 2019](#) and [UNICEF Brand Toolkit 2012](#). They must be language-edited / proof-read by a native speaker.

All reports will be in Microsoft Office Word format while all presentations will be in Microsoft Office PowerPoint. No PDF or hard copy will be submitted by the evaluation team. No page limit is set but all deliverables should be of the necessary length but not longer, and should be reader friendly. The use of bullet points, tables, graphs, photos, and other visualization methods is encouraged. The use of annexes is required for the evaluation tools, for all secondary information that is not directly related to the evaluation findings, as well as for any long technical documentation intended to a specific audience. PowerPoint presentations must include notes below each slide to make them easy to understand for people who could not attend the meeting.

All data collected, documentation gathered, and photos/videos taken and analyses produced for the purpose of the evaluation are to be made available to UNICEF in the appropriate format. Graphs and maps must be in editable format for layout purposes. All key deliverables will be made available on the UNICEF public website and widely disseminated to all target audience

6- Evaluation management and quality assurance

The evaluation will be conducted by an external evaluation team recruited by UNICEF EAPRO. The evaluation team will be led by a team leader, whose role will be to:

- Oversee and manage the other team members;
- Manage communications with UNICEF;
- Ensure adequate workspace, work equipment, accommodation, international and in-country travel, food, security and insurance arrangement to all evaluation team members;
- Orient and train team members where applicable;
- Organize and facilitate key meetings and workshops;
- Ensure that deadlines are met and all draft and final evaluation products and deliverables are of the required quality as per UNICEF standards (format and content).
- If necessary, the team leader will mobilize a dedicated quality assurance person for additional quality review.
- All draft and final deliverables submitted by the evaluation team leader will be accompanied by a detailed comment matrix describing whether and how earlier comments received have

been incorporated, and when they have not been fully included, providing an appropriate justification.

The evaluation team will operate under the guidance and supervision of an evaluation management team led by the Regional Evaluation Advisor and the Regional Evaluation Consultant with support from the SEA-PLM Project Manager and the Regional Education Advisor/Specialist and the Reference Group. The evaluation management team will be responsible for the contractual aspects, day-to-day oversight and management of the evaluation as well as evaluation budget. They will facilitate the communications with the country offices, the reference group members (see below), and other relevant stakeholders. They will be also responsible for the quality of the evaluation, and provide the first round of comments to the evaluation team before submission of the revised draft to the reference group. They will check whether the findings and conclusions from the evaluation are relevant and recommendations are implementable, and propose improvements to the recommendations. They will approve all deliverables and payments. In addition, the team will contribute for dissemination of the evaluation findings and to follow-up on the evaluation recommendations with a management response.

The role of the country office Education and Evaluation teams will be the following:

- Designate a focal point for supporting the evaluation,
- Provide the evaluation team with all information, resource documents and contacts necessary for the evaluation,
- Facilitate the communication and coordination between the evaluation team and UNICEF's implementing partners in-country,
- Organise / facilitate logistics, security, meetings and workshops; any related costs will be part of the regional evaluation and not from county office budgets,
- Provide comments on the key deliverables to minimize factual errors, misinterpretations, and omissions

UNICEF will provide quality assurance on all evaluation tools and documents based on the [UNEG](#)'s and [UNICEF](#)'s norms, standards, and guidelines as well as on other best practices related to Programme evaluations. Once approved, the final evaluation report will be submitted to the UNICEF's global evaluation reports oversight system ([GEROS](#)) for an independent quality rating. The report and the review will be made available on the UNICEF Internet website, in compliance with the commitment for transparency of evaluation findings.

Evaluation Reference Group

A reference group will be established with the following people and led by the evaluation management team.

- o Regional Education Advisor, EAPRO, UNICEF
- o Education Specialist, EAPRO, UNICEF
- o Regional Evaluation Advisor, EAPRO, UNICEF
- o Chiefs of Education section, UNICEF COs in EAPR (3 from case study countries, 2 from other countries)
- o Education Specialist, HQ, UNICEF

- Representatives from development partners (donors: Korea, Australia, NZ, ASEAN, JPE, WB, ADB, OECD)
- SEA-PLM Secretariat representatives
- Three (3) SEA-PLM 2019 participating countries
- Two (2) SEA-PLM 2019 non-participating countries

Additional members may be identified during inception phase.

The reference group will have the following roles.

- Generally, advise the evaluation management team on various aspects of the evaluation and help this team make decisions
- Contribute to the preparation and design of the evaluation
- Provide feedback and comments on the second draft of the inception report and on the technical quality of the work of the consultants
- Assist in identifying internal and external stakeholders to be consulted during the evaluation process
- Participate in review meetings organized by the evaluation management team
- Provide comments and substantive feedback from a technical point of view to ensure the quality of the second draft and final evaluation reports
- Propose improvements/inputs to the preliminary recommendations
- Play a key role in learning and knowledge sharing from the evaluation results
- Contribute to disseminate the findings of the evaluation
- Advise on the management response to the evaluation, and follow up when appropriate

7- Location and Duration

The indicative start date of the assignment will be 15 December 2019. The proposed institutional contract will have a duration of one year months, with all work being carried out between December 2019 and November 2020. The consultant will work remotely from home, with travel as agreed and two travels to Bangkok to present the inception report and the final report recommendations.

8- Qualification requirements and required experience

The core evaluation team may be comprised of 3 experts, while UNICEF is flexible in the team's composition so long as the evaluation work is of high quality. A gender balanced and culturally diverse team composition is strongly encouraged.

The consulting company must provide details of qualification and a work profile of the team leader who should meet the following required minimum conditions:

- Have proven experience (minimum 8 years of experience) in the use of participatory, qualitative and quantitative evaluation/analytic methodologies with experience in conducting UN evaluations;

- Have a minimum of MA/MSc level in Education, Social Sciences, Public Policy Analysis or other relevant fields related to Education. Experience of at least 5 years of work in the consultation domain;
- Evidence/proof of experience in a similar field;
- Strong communication (written and oral) and networking skills;
- Have proven competencies in research, facilitation skills, good communication, excellent English language and writing skills.
- Knowledge of Education in the East Asia and Pacific region is an asset.

Other team members should have the following qualifications:

- Advanced degree in related field with relevant work experience.
- Minimum 10-15 years relevant professional experience, ideally in education and large-scale assessment.
- Previous experience working in issues relating to education, large-scale assessment and/or related fields.
- Strong experience in research, program management, monitoring and reporting.
- Excellent interpersonal skills and previous experience communicating with partners at different levels.
- Demonstrated ability in work planning and report preparation.
- Demonstrated skills in professional high-quality writing in English.
- Experience working in the East Asia Pacific region highly desirable.
- Excellent communication skills in English.
- Experience working for a United Nations organization an asset.

Annex B - Learning and equity across Southeast Asia

As can be seen in Figure 7 and Figure 8, average proficiency scores in Singapore (in both Maths and Science) have sat between the High and Advanced International Benchmarks since 1995. Thailand's average proficiency scores have sat below or near the Intermediate International Benchmark for both domains and Malaysia's average proficiency scores have declined and then improved both above and below the Intermediate International benchmark, since 1999.

Figure 7 Trends in Grade 8 Maths Proficiency, TIMMS

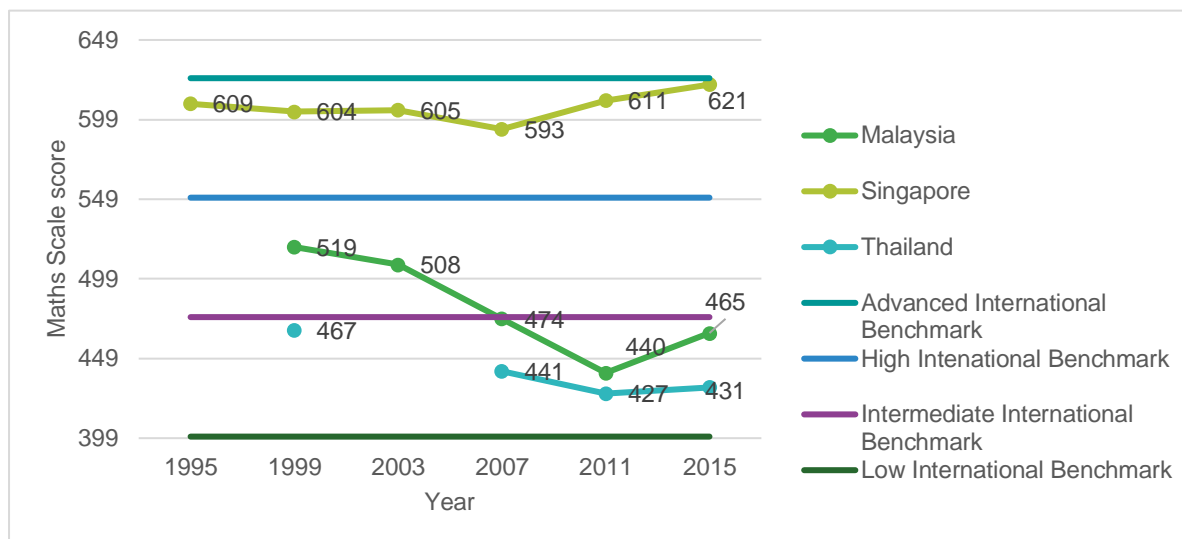
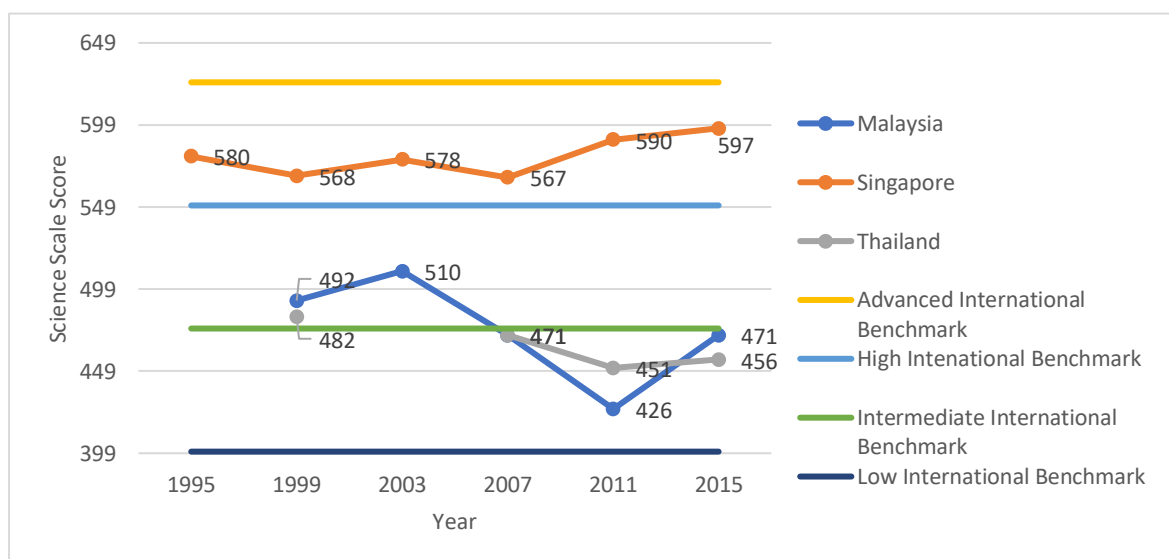


Figure 8 Trends in Grade 8 Science Proficiency, TIMMS



Gender differences in proficiency can be observed in Malaysia, Singapore and Thailand (TIMMS 2015). Girls perform higher than boys in each country in Maths and perform higher in Science in Malaysia and Thailand, while boys perform higher in Science in Singapore.

Figure 9 Average Proficiency in Maths, by Gender (TIMMS 2015)

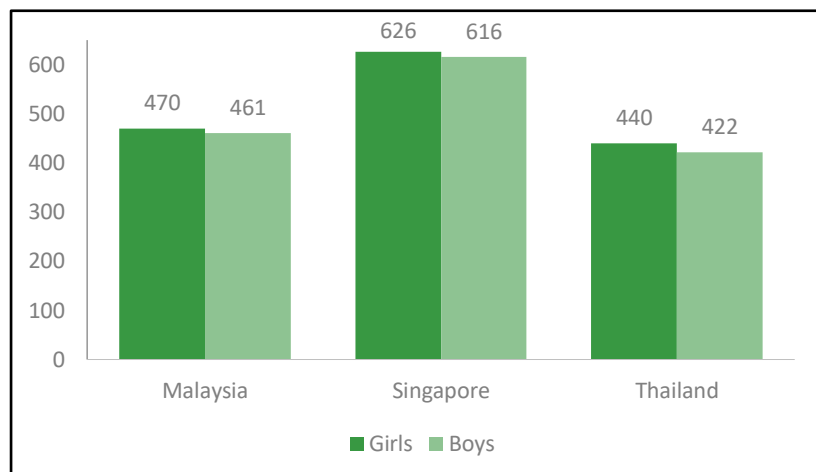
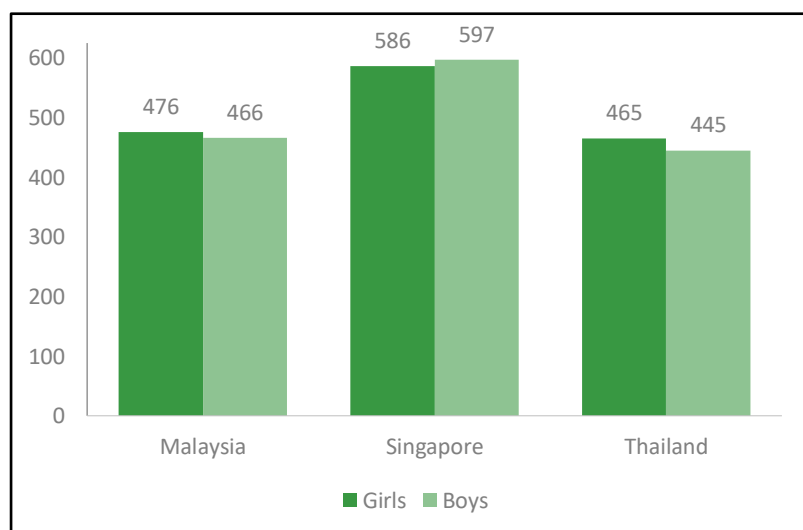


Figure 10 Average Proficiency in Science, by Gender (TIMMS 2015)



The higher performance of girls in maths is also observable in Vietnam, reflected in the Young Lives Secondary School wave one and wave two survey results. However, larger inequalities exist between districts, ethnic group and, for some measures, urban and rural divide (Iyer, Azubuike, & Rolleston, 2017, p. 31).¹¹ This highlights that forms of inequality beyond gender are also important to understand how disadvantage takes hold across contexts.

¹¹ This was also observable for English and Transferable Skills performance and for wave 1 and wave 2.

Annex C - Education System Quality Indicators in SEA-PLM participating countries graphs

Figure 11 Primary school age population (in millions) 2019¹²

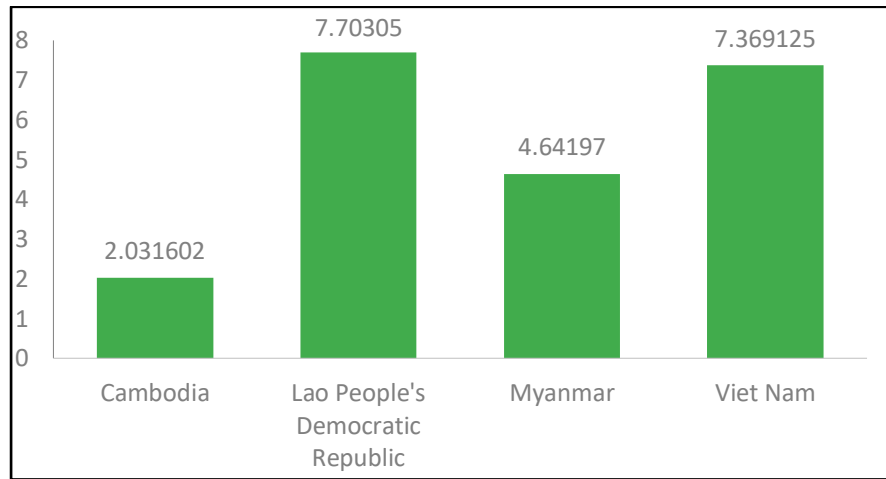
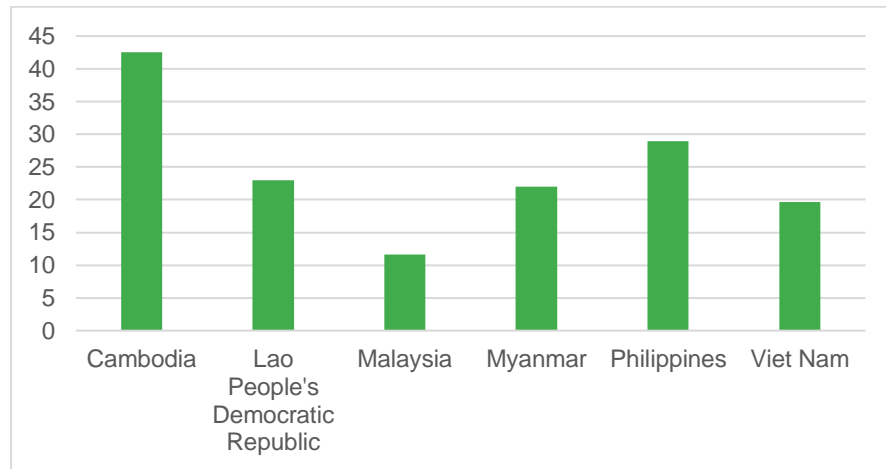


Figure 12 Pupil-teacher ratio in primary education¹³



¹² Source UIS

¹³ Source UIS

Figure 13 Percentage of trained primary teachers¹⁴

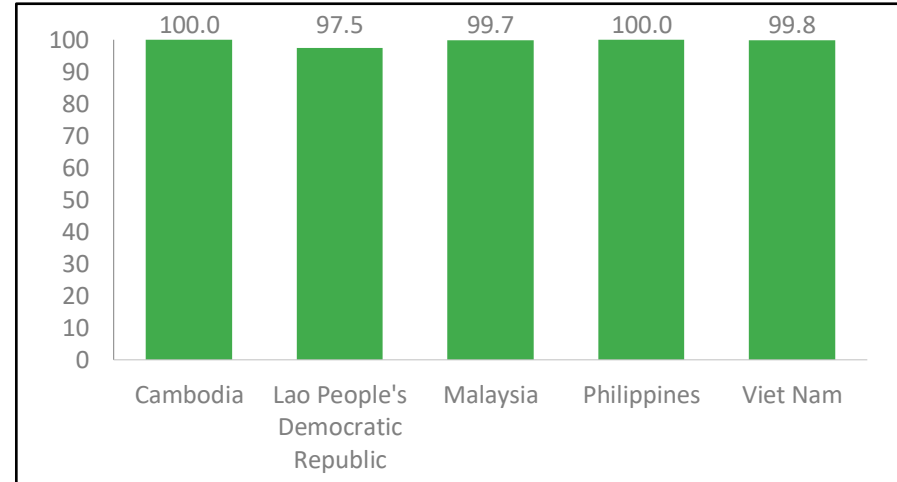
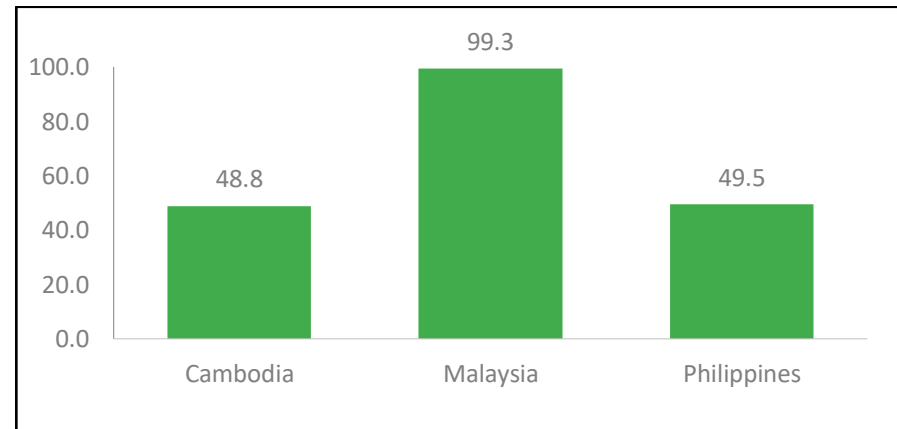


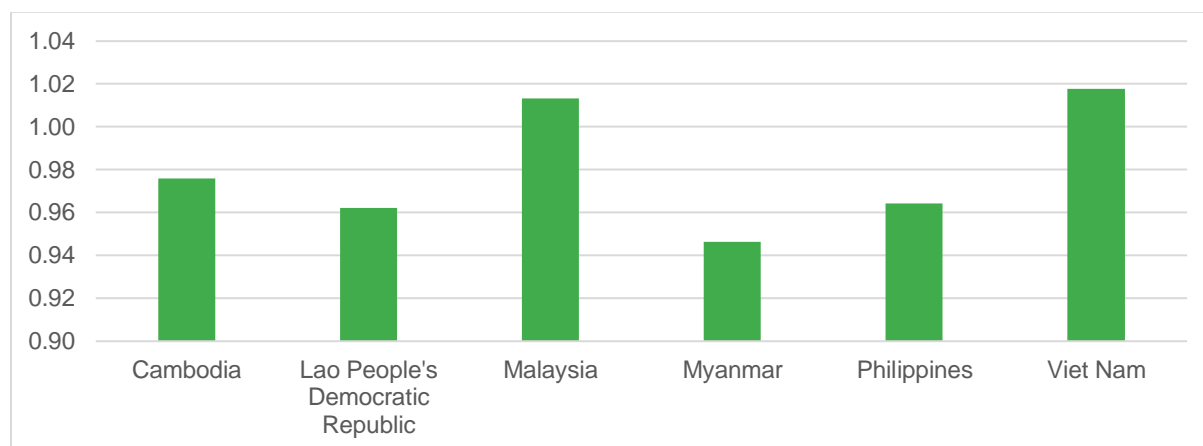
Figure 14 Proportion of primary schools with basic handwashing facilities¹⁵



¹⁴ Source UIS

¹⁵ Source UIS

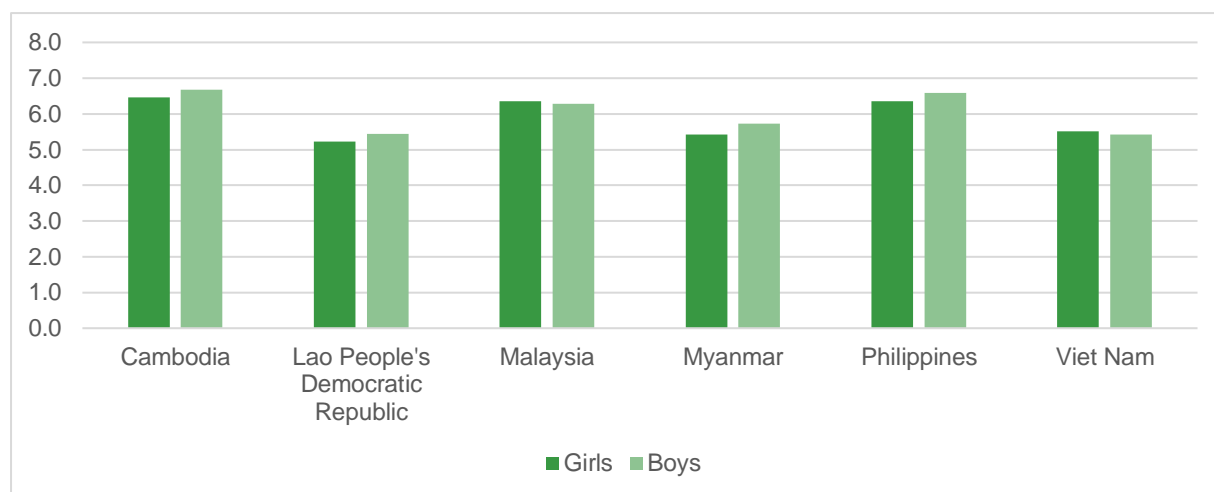
Figure 15 Primary Gross enrolment ratio, Primary, Gender parity Index¹⁶



Access, retention, survival and equity in SEA-PLM participating countries

Across the six SEA-PLM participating countries, primary school life expectancy tends to be higher for boys than for girls (Figure 16), yet survival rates to the last grade of primary education is higher for girls (Figure 17). This indicates a high rate of repetition for boys, which can be observed in Figure 18.

Figure 16 Primary school life expectancy, by gender 2017¹⁷



¹⁶ Source UIS. GPI is a socioeconomic index usually designed to measure the relative access to education of males and females.

¹⁷ Source UIS

Figure 17 Survival rate to the last grade of primary education, by gender¹⁸

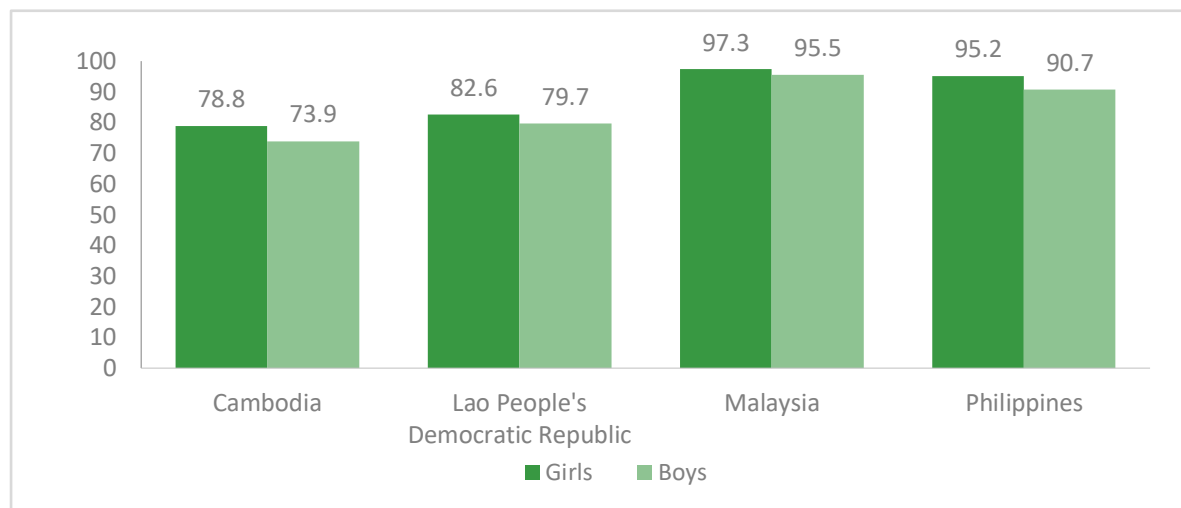
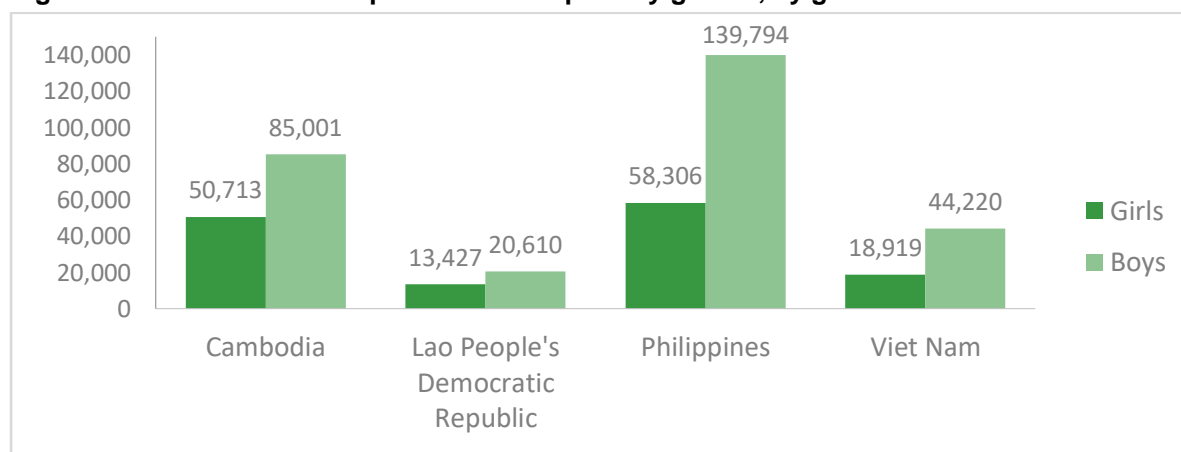
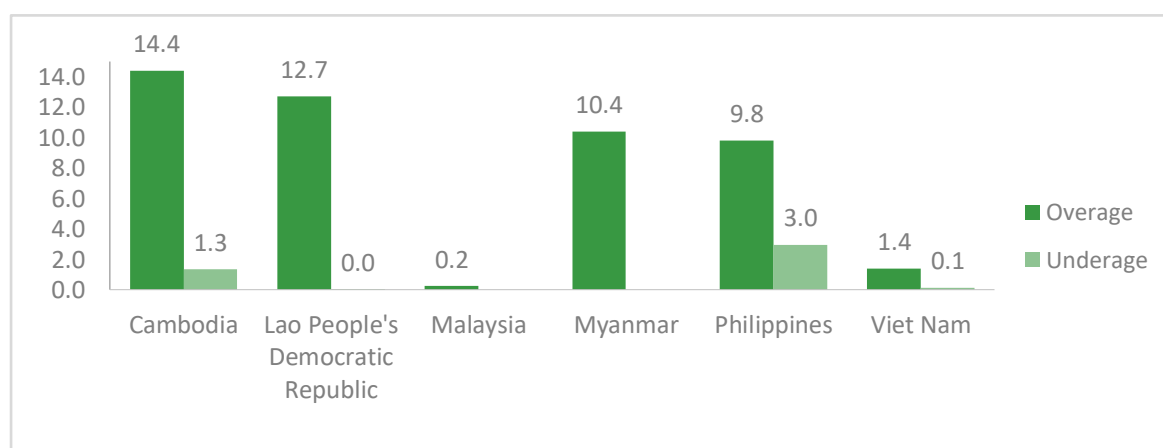


Figure 18 Number of repeaters across primary grades, by gender¹⁹



In Myanmar, Philippines, Lao PDR and Cambodia more than ten percent of the students enrolled in primary school are overage for their grade.

Figure 19 Percent of enrolled primary school students who are overage and underage²⁰



¹⁸ Source UIS

¹⁹ Source UIS

²⁰ Source UIS

Annex D - Partner support to education in SEA-PLM participating countries

In 2017 GPE's CEO Alice Albright visited Japan and Australia to meet with partners to discuss how best to collaborate to improve education in the Asia Pacific (Global Partnership for Education, 2020b). Cambodia has been a GPE partner country since 2006, and UNICEF and UNESCO are the Grant Agents (Global Partnership for Education, 2020a) and UNICEF is the Coordinating Agency. The Australian Department of Foreign Affairs and Trade (DFAT) invests in Cambodia, with one of the main objectives of Australia's aid being to better health and education outcomes (DFAT, 2020a). Only 8 percent of total ODA to Cambodia goes to the education sector and Japan and the Asian Development Bank contribute the most (OECD, 2020).

Lao has been a GPE partner country since 2009. The Coordinating Agencies include the DFAT, the European Union and UNICEF and the World Bank is the Grant Agent (Global Partnership for Education, 2020c). DFAT's Overseas Development Assistance (ODA) to Laos focuses on improving access to quality basic education for disadvantaged girls and boys. 18 percent of ODA to Lao goes to the education sector, with the Asian Development Bank and Japan being the largest donors (OECD, 2020). 39 percent of Malaysia's official ODA goes to the education sector, with Japan and the United Kingdom being the biggest donors (2017-2018) (OECD, 2020).

Myanmar has been a GPE partner country since 2018, with the Embassy of Finland and UNESCO filling the role of Coordinating Agency and the World Bank is the Grant Agent (Global Partnership for Education, 2020e). Education is the flagship of Australian aid to Myanmar, accounting for around 40 percent of bilateral development assistance in the country. Australia's investments focus on improving school learning environments, enhancing teaching and learning practices, supporting disadvantaged girls and boys to stay in school and strengthening the government's education policy development and oversight (DFAT, 2020b). The Department for International Development's (DFID) Myanmar UK Partnership for Education (MUPE) programme aims to improve the reach and quality of education services and help prepare the nation's youth with the skills they need for life after school. Importantly, the Assessment and Education Reform Support (AERS) component of MUPE is a five-year programme that will provide technical support to the Government of Myanmar (GOM) MOE to implement their national basic education assessment policy. This will include provision of training and other capacity development, policy analysis, communications and monitoring to ensure effective and accountable implementation of assessment reforms. Only 4 percent of aid is allocated to the education sector in Myanmar and Japan and the World Bank are the largest donors (OECD, 2020).

The Asian Development Bank supports the Philippines national flagship program in education and social protection and the US flagship initiatives "Basa Pilipinas" project to support an early grade reading program. Eight percent of total Overseas Development Assistance (ODA) was allocated to education in 2017-2018, with Japan and the United States contributing the most (OECD, 2020).

Vietnam has been a GPE partner since 2003, with UNICEF and the Coordinating Agency and the World Bank as the Grant Agent (Global Partnership for Education, 2020d). Eight percent of ODA went to education in 2017-2018, with Japan and the World Bank (International Development Association) being the largest contributors. However, on 1 July 2017, Vietnam officially graduated from the International Development Association, meaning that based on World Bank rules, Vietnam could no longer access concessional finance from IDA (OXFAM, 2019).

Annex E - Evaluation matrix

Evaluation Criteria	Evaluation question	Judgement criteria	Indicators	Research methodology	Source of information	Link to the ToC
Relevance	1. To what extent is SEA-PLM aligned to i) the national development policies and priorities of national education stakeholders and ii) UNICEF's national, regional and global objectives and intended impacts?	<p>1.1 Evidence of national development policies and priorities alignment with SEA-PLM objectives</p> <p>1.2 Evidence of UNICEF global, national and regional objectives aligning with SEA-PLM intended impacts</p>	<p>National stakeholder support for SEA-PLM</p> <p>Country and national institutional needs against programme delivery</p> <p>SEA-PLM compliments and builds on previous or existing initiatives in the region</p> <p>UNICEF policy and programme documentation aligns with SEA-PLM ToC</p> <p>There is sufficient domain coverage, international and regional data and political will to support a common regional and global metric</p>	Desk research, QCA	Country policies, UNICEF national, regional and global documentation, Programme documentation, KIIs, assessment frameworks and documents for other regional initiatives	Context, inputs and activities
	2. What are the programming gaps or unaddressed needs?	2.1 Evidence of any activities do not work within the context/s	Programme delivery is not on track against work plans	Desk research, QCA	Programme documentation, KIIs	Context, inputs, activities

	<p>What could be done better?</p>	<p>2.2 Evidence that needs (regional, national or sub-group), as identified by stakeholders, are not covered by SEA-PLM</p>	<p>When delivery is on track, additional needs are identified by stakeholders</p> <p>Programme delivery does not cover the needs of specific stakeholders or beneficiaries</p> <p>Programme activities do not lead to expected outputs</p> <p>Stakeholders have the opportunities (resources, time, conducive environment) to engage in SEA-PLM activities</p> <p>Stakeholders have the motivation and incentives to participate in SEA-PLM activities</p> <p>The reach, fidelity and dose of capacity building activities are sufficient at national and regional levels</p>			<p>and outputs</p>
--	-----------------------------------	-----------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--	--------------------

<p><u>Effectiveness</u></p>	<p>3. To what extent have the objectives and expected outcomes of SEA-PLM been achieved or are likely to be achieved?</p>	<p>3.1 Evidence that policy makers are preparing to use SEA-PLM data to inform decisions</p> <p>3.2 Evidence of improved regional and national capacity to produce, analyse and disseminate data</p> <p>3.3 Evidence stakeholders are aligning national assessments with SEA-PLM</p> <p>3.4 Evidence that SEA-PLM is supported by national governments and is on the regional agenda</p>	<p>Data is on track to be available to policy makers</p> <p>Policy makers have sufficient capacity, incentives and political space to access and utilize data to inform decisions</p> <p>Policy makers are aware of when the data will be released and how it can be used</p> <p>There is sufficient common ground between countries to find agreement on the analysis and dissemination of SEA-PLM results</p> <p>The data generated by countries is of sufficient quality to support a regional common learning metric</p> <p>There is sufficient national capacity (technical capabilities, political</p>	<p>QCA</p>	<p>KIIs</p>	<p>Outcomes and impact</p>
------------------------------------	---------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-------------	----------------------------

			will, resources) to analyse available data			
	4. What have been the major factors influencing the achievement or non-achievement of SEA-PLM objectives and activities? What were the enabling factors, barriers and bottlenecks?	4.1 Strength of evidence linking activities to outputs to outcomes 4.2 Evidence of triangulated perceptions regarding enabling factors, barriers and bottlenecks from key stakeholders	Current programme implementation status against workplans Progress against activities and outputs listed in the developed ToC Stakeholder perspectives on programme activities, outputs and outcomes Secretariat leverage within SEA countries and influence re SEA-PLM participation and SEA-PLM activities	Desk research, QCA	SEA-PLM programme documentation KIIIs	Activities, outputs and outcomes
	5. What can UNICEF and its partners do to ensure the objectives of SEA-PLM are met in the future? What kind of initiatives should UNICEF prioritize at the country and regional?	5.1 Evidence of particularly useful mechanisms of change along the results chain		Desk research, QCA	SEA-PLM programme documentation KIIIs	Complete results chain

<p><u>Efficiency</u></p>	<p>6. What were the strengths and weaknesses of SEA-PLM management processes? How could management of SEA-PLM activities be improved? To what extent and how did the UNICEF Regional Office contribute to efficiency of SEA-PLM in education programming?</p>	<p>6.1 Evidence of triangulated stakeholder perspectives and programme documentation regarding activity and output efficiency</p>	<p>Planned against actual costs</p> <p>Inputs were managed to achieve planned outputs within the resource envelope</p> <p>Planned activities led to planned outputs</p>	<p>Desk Research, QCA</p>	<p>Cost data and unit prices</p> <p>SEA-PLM programme documentation</p> <p>Official documentation on partner engagement</p> <p>KIIs</p>	<p>Inputs to outputs</p>
<p><u>Likely Impact</u></p>	<p>7. How has SEA-PLM contributed to national education systems and assessment practices and policies and discourses thus far? And at the regional level, what positive and negative changes has SEA-PLM brought about?</p>	<p>7.1 Evidence of changes in the topics, focus and amount of education dialogue at the country and regional levels</p> <p>7.2 Evidence in changes to national assessment practices</p>	<p>The use of SEA-PLM data leads to the improvements of previous shortcomings in relation to relevance, equity and learning in the education system</p> <p>Stakeholders have the capacity and incentives to solve education sector issues</p> <p>ASEAN countries see the value of a common approach to regional assessment</p>	<p>Desk Research, QCA</p>	<p>SEA-PLM programme documentation</p> <p>Country engagement in regional and international activities</p> <p>Country education plans</p> <p>Education Sector Analysis reports</p> <p>KIIs</p>	<p>Outputs to outcomes</p>

			<p>and believe SEA-PLM is sustainable.</p> <p>Stakeholders have the opportunities (resources, time and a conducive environment) to solve education sector issues</p>			
<u>Sustainability</u>	8. To what extent can SEA-PLM activities, plans and strategies be fully integrated and implemented by the government (s), both technically and financially? To what extent are they likely to continue?	<p>8.1 Evidence of capacity expectations to implement future rounds being achievable within TA budgets</p> <p>8.2 Evidence of national technical capacity to undertake technical tasks</p> <p>8.3 Evidence of funding availability and SEA-PLM affordability for future rounds</p>	<p>Costs of participating in SEA-PLM against other education expenditure in country</p> <p>Turnover of staff in participating countries</p> <p>Technical and capacity building inputs are sufficient to result in the desired outputs</p> <p>The costs of SEA-PLM are sustainable during integration into regional systems and beyond</p> <p>Donors and participating countries have</p>	Desk Research, QCA	<p>SEA-PLM programme documentation</p> <p>Country education budgets</p> <p>KIIs</p>	Inputs to impact
	9. What are the key barriers and	9.1 Evidence countries do not plan to				

	bottlenecks towards achieving sustainability of SEA-PLM activities?	participate in the future	sufficient resources to invest in SEA-PLM and consider SEA-PLM Value for Money (VfM)			
	10. How can SEA-PLM attract other countries in Southeast Asia and better link other international and regional initiatives?	10.1 Evidence of ways to overcome barriers and bottlenecks from other regional experiences	N/A	Desk research, QCA	Literature review KIs	N/A
<u>Equity and gender equality</u>	11. To what extent is SEA-PLM conducive to supporting the most marginalized populations and genders (including those furthest left-behind)?	11.1 Evidence that equity and inclusion has been mainstreamed into programme activities 11.2 Evidence the data will be used to inform equity decisions	Items are checked for differential performance (DIFF analysis) by gender or location Policy makers are considering equity, inclusion and gender in their policy concerns SEA-PLM data is able to capture inequalities and countries know how to disaggregate analysis by social group	Desk research, QCA	Programme documents KIs Data analysis plans	Outputs to impact

Annex F - Logframe

Programme Impact	ASEAN and SEAMEO Member countries have an enhanced understanding of factors affecting learning achievements in primary education and take actions to ensure that all children achieve meaningful learning outcomes.		
Programme Outcome 1	Enhanced capacity to generate and analyse assessment data at regional, national and sub-national levels		
Outputs	Indicator	Assumptions/Risks	
Output 1.1: National officials have necessary capacity to conduct SEA/PLM surveys			
<p>Activity areas 2015- 2016:</p> <p>Assessment tools developed, tested, and finalised for use in national surveys</p> <p>Key staff trained at regional and national levels in item development, test design and administration</p>	<p>Assessment tools ready for use in main survey</p> <p>Key staff have required competences to carry out survey effectively</p>	<p>Process at country level includes key experts and decision-makers</p> <p>Trained staff will be utilised in actual survey exercise.</p>	
Output 1.2: SEA/PLM surveys meet expected standards of efficiency and effectiveness			
<p>Activity areas 2015-2016:</p> <p>Field testing of SEA/PLM instruments</p> <p>Activity areas 2017:</p> <p>National surveys completed</p>	<p>Instruments field tested in at least 6 countries (2 rounds of field testing planned)</p> <p>National surveys completed in at least 6 countries</p> <p>Final reports meet global standards for assessments</p>	<p>Funding is sufficient to ensure that national surveys are carried out efficiently and effectively.</p>	

Programme Impact	ASEAN and SEAMEO Member countries have an enhanced understanding of factors affecting learning achievements in primary education and take actions to ensure that all children achieve meaningful learning outcomes.	
Programme Outcome 2	Enhanced capacity to utilise assessment data for education improvement and more equitable learning outcomes at regional, national and sub-national levels	
Outputs	Indicator	Assumptions/Risks
Output 2.1: Future national sector policies and plans respond to SEA/PLM findings	SEA/PLM data cited in sector policies and plans	
<p>Activity areas 2015 - 2016:</p> <p>Country visits by SEAMEO, ACER and UNICEF to explain how SEA/PLM can contribute to enhanced learning and equity.</p> <p>Dissemination workshops at national level with key stakeholders</p> <p>Policy briefs developed based on SEA/PLM field trial findings</p>	<p># advocacy visits to potential SEA/PLM countries and visit reports as shared with SEA/PLM Steering Committee</p> <p># and quality of workshops</p> <p># and quality of policy briefs</p>	<p>High level political buy in at Ministerial level to accept, share and utilise survey findings</p> <p>Inclusion of all key stakeholders in the planning and dissemination processes</p> <p>Timing of SEA/PLM surveys in relation to education reform opportunities</p>
Output 2.2: Regional analyses on progress in education include data from SEA/PLM surveys	SEA/PLM data cited in regional analyses of education	
<p>Activity areas 2017:</p> <p>Regional meetings to discuss and compare SEA/PLM findings across Member Countries</p> <p>Regional analytical briefs developed.</p>	<p># of meetings and quality of recommendations</p> <p># of briefs developed and evidence of impact</p>	<p>Regional education co-ordination structure for Post 2015 remains strong and harmonious</p> <p>Regional analyses must not lead to negative competition and league tables that shame low performers but to collaborative approaches to enhancing learning for all children</p>

Programme Outcome 3	Enhanced ASEAN integration in the education sector in terms of common approaches to assessment and enhancement of learning, quality and equity	
Outputs	Indicator	Assumptions/Risks
Output 3.1: SEA/PLM endorsed as regional priority by most ASEAN and SEAMEO Member Countries	# of states agreeing to engage in SEA/PLM increases (Baseline. July 2015: 4)	
<p>Activity areas 2015 -2016:</p> <p>Regional SEA/PLM meetings hosted by SEAMES include advocacy around regional assessment</p> <p>Advocacy for SEA/PLM during SEAMES regular meetings and visits abroad</p> <p>Communication strategy for SEA/PLM, including brochure and web portal, developed and shared across all SEAMEO Member Countries</p> <p>Regional SEA/PLM country communication package</p> <p>Advocacy with ASEAN to ensure inclusion of SEA/PLM in ASEAN workplan</p> <p>Resource mobilisation for SEA/PLM through SEAMEO and ASEAN structures and through engagement of additional donors</p> <p>Activity areas 2016 - 2017</p> <p>Review of 2015 field trials shared at key meetings</p> <p>Formative external evaluation of SEA/PLM progress and impact</p>	<p># of regional meetings on SEA/PLM/Assessment</p> <p>#number of SEAMEO meetings and visits where SEA/PLM was presented</p> <p># of people using new portal and</p> <p>#of brochures distributed</p> <p># and quality of adapted country communication documents</p> <p>SEA/PLM cited in ASEAN plan</p> <p># of new donors</p> <p>Level of funding raised</p> <p>Level of national investment in SEA/PLM</p> <p>Key findings enhance survey design</p> <p>Key findings used to shape development of SEA/PLM</p>	<p>SEAMES advocacy with member states will be critical</p> <p>Some countries may resist due to commitments to other surveys</p> <p>ASEAN yet to engage with SEA/PLM as regional priority.</p> <p>Need ensure SEA/PLM is not seen to be donor driven</p> <p>Donor resistant to investing in less tangible areas</p>

Annex G - Refining the ToC

The ToC workshop was undertaken with the SEA-PLM Reference Group on 2nd June 2020. The content included:

1. Introduction to the evaluation team;
2. And overview of the purpose and scope of the evaluation
3. A summary of the evaluation methodology, including evaluation questions, the Evaluation Matrix and the data collection activities and instruments
4. An interactive session designed to elicit feedback on the reconstructed ToC; and
5. The evaluation outputs and engagement.

Using Mentimeter online software, the Reference Group were provided with a set of statements and were asked to state the extent to which they agree with the statements on a scale of 1-5. In order to add qualitative information to the quantitative data, discussion on the key areas of inputs, activities, intermediate outcomes, longer terms outcomes and Impact were undertaken, and participants were invited to submit further written perspectives to the evaluation team.

The feedback provided by the Reference Group through written feedback, quantitative mentimetre data and qualitative discussion has been integrated with the reconstructed ToC as described in the draft inception report and is described and presented below.

Impact

Various documents refer to the impact of SEA-PLM, but often statements combine outcomes and impact statements. For example the SEA-PLM (2015) Background, Rationale and Programme Logic states that the desired programme impact is: *“Countries from the Association of Southeast Asian Nations (ASEAN) and the Southeast Asian Ministers of Education Organization (SEAMEO) have an enhanced understanding of factors affecting learning achievements in primary education and take actions to ensure that all children achieve meaningful learning outcomes”* (SEA-PLM, 2015, p. 4).

The same document refers to a vision, stating *“SEA/PLM will contribute towards improving and redefining learning outcomes by providing a regional culturally appropriate metrics and thereby towards a more equitable and meaningful education for all children across the region”*.

In the 2019 SEA-PLM Communication Strategy the overall goal of SEA-PLM is expressed as *“by 2025, all Southeast Asia countries have regional standardized quality measures to assess learning outcomes for grade 5 students in the domains of reading, writing, math and global citizenship and consistently use these to inform improvement within the education sector”* (SEA-PLM SEAMEO UNICEF, 2019b, p. 16).

Both of these statements combine impact, outcomes and process. However, taken together, the three phrases taken from the desk study confirm that the desired impact of SEA-PLM is ‘improved, and more equitable, relevant student learning outcomes across Southeast Asia’.

During the ToC workshop, the Reference Group supported an impact statement that explicitly references and is more closely aligned to Sustainable Development Goal 4 (SDG4). For this reason, the impact statement was adapted to:

More inclusive, equitable quality education for all in Southeast Asia, as per SDG4.

Longer term and intermediate Outcomes

The ToC mapping included in the Communications Strategy (2019) indicates that the three goals of SEA-PLM are:

- Enhanced capacity to generate and analyse assessment data at regional, national and sub-national levels
- Meaningful utilization of assessment data for education improvement and more equitable learning outcomes; and
- Enhanced ASEAN integration in terms of approaches to assessment.

When SEA-PLM was initiated in 2012, its key objectives were to develop a regional learning metric to allow for a common approach to assessing learning outcomes at age 10 or Grade 5, which countries would adopt over time across the Southeast Asia region. Reference Group members also confirmed that over time discussions have also included developing additional measures over time and across additional grade levels.

It was assumed that the initiative would take shape through the combination of SEAMEO's political leadership, technical input from an expert contractor (ACER), and programmatic support of UNICEF, in conjunction with SEAMEO member countries (Ahmad, 2015, p. 4).

Similarly, the SEA-PLM Assessment Framework (2017) states that the 'SEA-PLM initiative has three core goals:

- To provide policy makers with relevant, sound and comparable data on contextual and learning outcomes that can directly inform local education policy development.
- To develop indicators of educational outcomes that enable meaningful comparisons of quality.
- To enhance the existing capacities of participating countries to design data collection activities that will assist all aspects of the policy cycle: to develop and implement a reliable, valid and rigorous survey-based assessment and reporting program; and to appropriately analyse, interpret and disseminate assessment data with a view to informing education policy through relevant evidence.' (UNICEF and SEAMEO, 2019a, p. 5)

Other desired outcomes include utilizing the data towards policy priorities. These include informing teaching practices, informing curriculum scope and sequence and informing community based programs (Spink, 2019). The intended use for the data is explicated in Spink (2019); *"Assessment results must help us understand where children are at, so we can best understand where we want children to move to next. Regional and global reporting requires us to compare but we must also be able to understand if we are to make a change."*

The desire for data to be used for decision-making is also reflected in UNICEF's global Education Strategy 2019-2030. The Strategy highlights systems strengthening in partnership with governments and also that *'the generation and use of data and evidence will be (further) enhanced, particularly related to levels of learning, to generate a better understanding of the children being left behind, and the effectiveness of education systems in meeting the learning needs of every child'*. (UNICEF, n.d.)

These goals and outcomes expressed within the documentation cover outputs, intermediate outcomes and longer-term outcomes. In order to use the ToC for evaluation purposes, the logical results chain must link longer term outcomes to impact and intermediate outcomes to longer term outcomes etc. For this reason, it is important that the sequencing of the logical chain follows the theory of how change occurs. For example, to link longer term outcomes to impact, the longer-term outcome would need to be able to be directly linked to how improved, and more equitable, relevant student learning outcomes across Southeast Asia would take place. *"Enhanced ASEAN integration in terms of approaches to assessment"* is not an outcome that can logically directly lead to improvements in learning outcomes. For this reason, the outcomes and goals referred to in

SEA-PLM documentation have been reordered in the reconstruction, to take account for the need for logical sequencing along the results chain.

The longer-term outcomes of SEA-PLM, as expressed in the documentation, and the mechanism by which impact can take place is 'sustainable, effective, efficient, responsive and more equitable education systems at the regional and national levels, able to meet the needs of all children, including those left behind across Southeast Asia.

Reference Group feedback indicated that this Longer term outcome is too long and should be simplified. For this reason, the Longer term outcome was adapted to:

A stronger education system, able to meet the needs of all children

Many of the objectives referred to in the documentation are theorised to lead to the longer-term outcome of a stronger, more equitable, education system. In the documentation these outcomes are referred to as:

1. Policy makers use data to meaningfully support decision-making;
2. Enhanced capacity to generate and analyse assessment data at regional, national and sub-national levels;
3. Sustainable regional integration of relevant large-scale assessment and outcomes;
4. Enhanced ASEAN integration in terms of assessment; and
5. SEA-PLM is on the regional agenda and national government support is re-enforced.

When asked the extent to which they agreed with each of these outcomes being SEA-PLM outcomes (scale of 1-5), participants at the inception workshop agreed with outcomes (as listed above) one (average 3.8), strongly agreed with outcome two (4.5) agreed with outcome 3 (3.3) and agreed with outcome 5 (3.4). Participants had mixed views on the extent to which SEA-PLM aims to enhance ASEAN integration in terms of assessment.

Based on discussions and further written comments on outcome four it was clarified that it would be more accurate to refer to support for learning assessment policies and practices at the regional and national levels, rather than integration in terms of assessment.

Importantly, the issues of accountability and political will to create an enabling environment for data to be used, were raised within the workshop and in subsequent feedback.

Governance for SEA-PLM was also highlighted as missing in the programme outcomes.

The intermediate outcomes were refined to:

1. Policy makers use data to meaningfully support decision-making;
2. Enhanced capacity to generate and analyse assessment data at regional, national and sub-national levels;
3. Sustainable regional integration and governance of relevant large-scale assessment;
4. Support for learning assessment policies and practices at the regional and national levels; and
5. Cross border dialogue, accountability, exchange and action facilitated from regional to school levels to improve learning.

Outputs

There are many specific outputs within the SEA-PLM programme that span the technical, political, logistical and partnership aspects of the programme. Within the ToC the broad and most important outputs have been identified, as referenced in the SEA-PLM conference paper (Ahmad, 2015), technical reports (SEA-PLM, 2019a), work plans (SEA-PLM SEAMEO UNICEF, 2019a and ASEAN Secretariat, 2016), communication strategy documentation (SEA-PLM SEAMEO UNICEF, 2019b), program rationale and logic documents (SEA-PLM, 2015), ToR (UNICEF EAPRO and UNICEF, 2019) and the SEA-PLM Assessment Framework (UNICEF and SEAMEO, 2019a).

These broad outputs identified in the desk study are:

1. Reliable, valid, relevant and rigorous learning and contextual data;
2. A regional common learning metric;
3. SEA-PLM alignment to the global UIS learning outcomes scale;
4. Regional and national workshops and Learning and Development (L&D) products;
5. Strategies for analysing and disseminating data;
6. Common tools, protocols and standards to generate and utilize assessment data across ASEAN countries; and
7. Funding agreements with donors and participant countries.

Reference Group feedback highlighted that the original focus on a regional common learning metric was too narrow and this output should be changed to 'sustainable and credible model for long-term assessment' and that a common metric at the regional level should be added to the output of alignment to a global UIS learning outcomes scale. Coordination and planning was identified as a missing output so this was added to the regional and national workshop output description. The importance of embedding the common tools, protocols and standards at national levels was identified as missing and so was added to output 6.

Activities

SEA-PLM ToC mapping documentation identifies four main areas of activities to achieve SEA-PLM goals, including:

- **Capacity building:** to improve procedures and capacities at individual and institutional levels for completing standardized assessment of grade 5 learner achievement
- **Data production:** to improve the common design of tools, protocols and standards to generate and utilize assessment data across ASEAN Member States
- **Information use:** so, data can be analysed, understood, and used by skilled decision-makers who take and promote evidence-based actions to improve education policy and curricula
- **Regional integration:** to ensure sustainable regional integration of large scale assessment to monitor achievements towards the ASEAN Community Blueprints and SDG 4.

In many ways, these activities articulate four groups of activities and then describe how they lead to outputs and outcomes. Therefore, in reviewing the range of SEA-PLM regional documentation, the main activities were identified. These activities are presented in Table

6. The information from the desk study was used to categorise and articulate the broad categories of activities for the ToC, including:

1. Field studies and main survey data collection;
2. Assessment framework development, item development, sampling, analysis and reporting;
3. Technical capacity building activities;
4. Develop and disseminate communications strategy;
5. Regional endorsement activities;
6. Fundraising with donors and countries.

Reference Group feedback through Mentimetre software indicated strong agreement with each of the above activities (averaging between 3.9 and 4.7 out of 5). However, qualitative feedback highlighted the importance of programme governance and management as an activity in and of itself, and the implementation of the communications strategy was also identified as an important activity. These components were added to the programme activities in the ToC.

Table 6 SEA-PLM Activities

2012	2013	2014	2015	2016	2017	2018	2019	2020
Inception	Research paper on education curriculum s	Formal collaboration established between SEAMES and UNICEF for the realization of SEA-PLM	Audit of all SEAMEO countries curricula completed	Vietnam, Malaysia and the Philippines confirm their participation in the field trail	First round of field trial results in Brunei, Cambodia and Laos	Planning and budgeting for main survey	First in country review of regional and national activity plan for 2020 prepared by SEA-PLM Secretariat and reviewed by the TAG	Regional report of cross-country findings launched
Working group established	Research paper on assessment programmes	ACER contracted to develop tools for regional assessment	Assessment Framework developed	Remaining participating countries trained several technical areas	Malaysia, Philippines and Vietnam complete their field trials	Completion of Governance Analysis	Regional Steering Committee in-person meeting to initiate reflection and strategize on a second version of regional and national activity plan for 2020.	Technical delivery: Creation of one provisional regional database with weighting.
Conceptualization	Regional Seminar held to agree key features and scope of SEA-PLM	Brunei, Laos and Cambodia signed up for field trail	Assessment Framework reviewed at regional workshop	Laos, Cambodia and Myanmar complete the field trial		TOR and Contracts for implementation of main survey at national and regional levels	In-country consultation on national activity plan for 2020	One regional technical meeting or webinar to share psychometric conclusions
		Formal launching ceremony with high-level representation and endorsement of national governance structure	Country visits from SEAMES and EAPRO to meet with High Officials to discuss SEA-PLM	SEA-PLM is included in ASEAN 2016-2020 Education Work Plan		Regional Consultation on next steps for SEA-PLM	Final validation of regional and national activity plans for 2020 comprising work plan, timeline, and budget	Technical support: one psychometric report
			Regional item development workshop			Main Survey Orientation and Sampling workshop held in 6 countries	New communication strategy, including newsletter diffusion, publication edition and first step new branding and program design for 2020	Technical support: one competency scale by domain

2012	2013	2014	2015	2016	2017	2018	2019	2020
			Regional Orientation workshop (up-start of field trails)			Code of Practice drafted	Implementation of the regional work plan and nationals' work plan in countries	Technical support: one analysis plan
			Myanmar confirms their participation in the field trial			National Sampling design completed in 6 countries	Regional report of cross-country findings launch (SEA-PLM SEAMEO UNICEF , 2019a)	Technical support: one provisional outline for the main regional report
			Brunei D completes their field trial			Key findings from the Governance Analysis presented- Regional SEA-PLM Consultation		Technical support: one data analysis user manual
						Accuracy of life skills learning assessment workshop		Technical support: one methodological report.

Inputs

The inputs to SEA-PLM fit into four broad areas:

- 1. Technical Assistance;**
- 2. Data collection costs;**
- 3. Regional and country governmental human resources; and**
- 4. UNICEF coordination.**

The financial inputs for SEA-PLM are complex and not easily traced. However, several documents made available for the Desk Study, provided some information on historical costs and future budgetary figures.

Technical support contracted through the UNICEF Long Term Agreement (LTA) totalled US \$2,127,623.20 between February 2017 and December 2019 (UNICEF, 2019). Overall institutional contracts established to date total just over \$5 million (UNICEF, 2019).

Data collection costs are not easily identifiable as discreet costs; however, documentation indicates that country implementation for field trials in 2018 totalled US \$1.66 million across Cambodia, Malaysia, Vietnam, Laos, Myanmar and the Philippines. This doesn't include technical support for field trails or other institutional contracts associated with field trails.

Regional costs for 2020 are currently costed at US \$715,176 and country costs are US \$727,794. In addition, Secretariat staff costs (not counting existing staff) are estimated at \$1.5 million over 2019-2021.

Of course, in addition to the direct costs associated with implementing SEA-PLM, human resources make up a large contribution of inputs from across various partners. For example, according to documentation from 2018, the SEAMES and UNICEF staff assigned to SEA-PLM included a Deputy Director, a Finance Manager, a Programme Officer, a Programme Coordinator, a Programme Assistant, a Regional Education Advisor, Education Specialist, a Senior Expert and a Programme Manager (UNICEF EAPRO).

According to the documentation, the funding gap for 2020 is significant, at approximately 75 percent of the regional and national costs (not accounting for any further institutional contracts) (Marivin, 2019).²¹

Reference Group members all indicated support to strong support for each of the SEA-PLM inputs, ranging from an average strength of agreement from 4.4 to 4.7 (out of 5).

Qualitative Reference Group provided more nuanced feedback. For example, one stakeholder stated that outcomes from early rounds of SEA-PLM, would become future inputs to SEA-PLM. This highlighted the importance of understanding the cyclic nature of SEA-PLM.

The multiple cycles of SEA-PLM assessments will build towards the longer-term goals as identified in the ToC. Each cycle is intended to contribute to learning and evidence, feeding into the next cycle and the overall programme is intended to adjust and evolve. The strength of outcomes in early rounds will input into future rounds (though are not 'inputs' for the purposes of a ToC).

²¹ Based on the technical and financial proposition of the contractor available at this time and before annual internal call for funding within UNICEF.

In the future outputs and activities will be subject to changes brought about by learning throughout the implementation of multiple cycles of SEA-PLM and necessary adjustments from each cycle are intended to be made.

For this reason, an arrow feeding from outcomes back to activities was added to the ToC to illustrate the cyclic nature of SEA-PLM as it moves through progressive rounds.

Annex H - Data collection tools

1. Introduction and explanation of the evaluation
2. Informed consent information
3. Confirmed or not confirmed consent [if not end interview]

Table 7 Interview items by stakeholder group

<i>[Questions to be tailored to the specific respondent.]</i>	UNICEF	Regional governance	National governance	ACER	TAG
Relevance					
Can you please tell me about your role and how you've come to be involved in SEA-PLM?	X	X	X	X	X
How did you first come to know about SEA-PLM? When was this?	X	X	X	X	X
What were your first impressions about the need for SEA-PLM?	X	X	X	X	X
Have our views changed since that time?	X	X	X	X	X
Can you tell me more about your institution/unit and its goals and aims?	X	X	X	X	X
Do you think these goals and aims align with the aims of SEA-PLM?	X	X	X	X	X
Is programme delivery of track at the moment? [if no] What are the main barriers to keeping to SEA-PLM workplans?	X	X	X	X	X
Are there any activities that aren't included in SEA-PLM that you think should be?	X	X	X	X	X
Efficiency					
What SEA-PLM activities have you been involved in?	X	X	X	X	X
Are the activities and outputs of SEA-PLM achievable? [If no]					
Probes: What have been the challenges? Can you give me an example of a time when it was challenging to achieve the activities or the outputs for SEA-PLM? How did you solve the problem?	X	X	X	X	X
Do you feel the technical support within SEA-PLM is provided when needed and to a sufficient standard? [Prompt for information]	X	X	X	X	X
Do you feel the financial support provided by UNICEF has been provided when needed?	X	X	X		

<i>[Questions to be tailored to the specific respondent.]</i>	UNICEF	Regional governance	National governance	ACER	TAG
Is the financial support sufficient to cover agreed SEA-PLM activities?	X	X	X		X
Do you feel UNICEF’s technical support is sufficient for participating countries and delivered on time?	X	X	X	X	X
What are some of the challenges with providing technical resources within SEA-PLM?	X			X	X
What are some of the challenges with managing financial resources to support SEA-PLM?	X				
What are your thoughts on the new governance structure for SEA-PLM?	X	X	X	X	X
Do you think there are ways UNICEF, regional government, national governments and contractors could work better together?	X	X	X	X	X
What stage is the country/region [depending on participant] on track to have the data from the main study clean and ready for analysis and dissemination?	X	X	X	X	X
Can you tell me more about the development of the common learning metric? Has this process been completed yet? Where there any challenges?	X	X	X	X	X
Can you tell me more about alignment of SEA-PLM to the UIS global learning outcome scale?	X			X	X
Have you participated in any technical or L&D workshops? [if so] Which ones?	X	X	X		X
What was your experience of the workshop/s?	X	X	X		X
What did you learn as a result of the workshop/s?	X	X	X		X
Have you applied what you’ve learnt in any other aspects of your work?	X	X	X		
In [country] have any of the approaches taken in SEA-PLM influenced the approach to national assessment?	X	X	X		
Can you tell me about the common tools and protocols that have been developed as part of SEA-PLM? Probe: Are they useful? Where are they available? Can you give me an example of when you’ve used them?	X	X	X		X
There has been a range of attempts to fundraise for SEA-PLM, what has been successful? What hasn’t worked? What has the feedback from potential funders been?	X	X			

<i>[Questions to be tailored to the specific respondent.]</i>	UNICEF	Regional governance	National governance	ACER	TAG
SEA-PLM aims to raise funds with donors and from participating countries. How does [country] currently contribute to the costs of participating in SEA-PLM? What do you think contributions will look like in the future?	X	X	X		
<i>Effectiveness</i>					
What are the main policy questions you think SEA-PLM can provide information on?	X	X	X	X	X
What questions do you have, that you think SEA-PLM data will shed light on?	X	X	X		
Do you think there will be any challenges with disseminating the data and ensuring it is used?	X	X	X	X	X
What are some of the key decisions you intend to make, based on SEA-PLM results?	X	X	X		
Do you think that decisions in [country] are based on evidence? Probe: Why/why not? What enables or inhibits evidence-based policy making?	X	X	X		
Who will analyse the data if a policy maker or a minister has a specific question that SEA-PLM can answer? Probe: Will there be any challenges?	X	X	X		
The Secretariat seeks to influence SEA countries to participate in SEA-PLM. What kind of activities have you engaged in to achieve this? What kind of outcomes have you seen? What are the best challenging aspects of influencing SEA countries?	X	X			
<i>Likely impact</i>					
What do you think would be different in [country or region] if SEA-PLM didn't exist?	X	X	X	X	X
What do you think are the biggest challenges to achieving education quality for all in [country or region]?	X	X	X	X	X
How does SEA-PLM contribute to solving education sector issues in [country or region]?	X	X	X	X	X
What is the value-add of a common approach to assessment in SEA?	X	X	X	X	X
<i>Sustainability</i>					

<i>[Questions to be tailored to the specific respondent.]</i>	UNICEF	Regional governance	National governance	ACER	TAG
Do you think [country] will participate in future SEA-PLM rounds? Probe: Why or why not?	X	X	X	X	X
What technical support will countries need to participate in the next round of SEA-PLM?	X	X	X	X	X
Has there been much turnover of staff who have received technical training in [country or region]?	X	X	X	X	X
What are the plans for meeting the funding gap for SEA-PLM in 2019 and 2020?	X	X	X		
What are the plans for meeting the costs of SEA-PLM in future survey rounds?	X	X	X		
Is SEA-PLM value for money?	X	X	X		
<i>Equity and Gender items are mainstreamed into the other items</i>					

Annex I - Interview guide

1. These guidelines are not intended as questionnaires. It will not be possible to cover all issues in all categories with all individuals or groups. The evaluation team members will use their judgement and focus on areas which are likely to add most to the team's existing knowledge, while allowing interviewees and groups to highlight the issues that are most important to them.
2. The evaluators will formulate questions in a (non-technical) way that respondents can easily relate to, while generating evidence that is relevant to the evaluation questions that the evaluators have in mind.

Approach to Interviews

3. Interviews will be a major source of information for this evaluation. These will be a means to extract evidence, as well as to triangulate evidence drawn from other interviews and the document review and will form part of the consultative process.
4. A stakeholder analysis as presented in baseline report will inform the selection of interviewees. Over the evaluation period the evaluation team aims to target a comprehensive range of stakeholders that fully represent all significant institutional, policy and beneficiary interests. The team will periodically review the list of those interviewed to ensure that any potential gaps are addressed and to prevent under-representation of key stakeholders.
5. All interviews will comply with the team's commitment to the respective evaluation ethics. (The work of the evaluation team will be guided by: OECD-DAC Evaluation Quality Standards for Development Evaluation; UNEG Norms, Standards, Ethical Guidelines and Code of Conduct for Evaluation in the UN System; the World Bank's principles and standards for evaluating global and regional partnership programs; ALNAP's Evaluation of Humanitarian Action Guide; the Sphere Handbook and Standards for Monitoring and Evaluation; and guidance on Ethical Research Involving Children.)
6. Interviews will be conducted in confidence and usually on a one-to-one or one-to-two basis (to enable note taking). Reports will not quote informants by name and will not include direct quotes where it could risk revealing the participant's identity or attribution without prior consent.
7. A protocol and standard format for recording interview notes is presented below. This will be used for all interviews and will ensure systematic recording of details, while allowing for flexibility in the specific questions asked. Interview notes will be written up, consolidated into an interview compendium and shared among team members via the internal team-only e-library. To respect interviewee confidentiality, the interview notes will be accessible only to team members. The compendium of interview notes will facilitate analysis across all interviews and will enable searches on key thematic terms, initiatives and so on. This will maximise the analytical potential of interviews and the possibilities for triangulation.

Interview template

Date of Interview:	
Location:	<i>Include whether remote or face-to-face</i>
Team members present:	

Notes by:		Date completed:	
Interviewees			
Name	m/f	Designation (position/unit/organisation):	Contact (email/phone)
<i>add rows for additional people.</i>		<i>Give sufficient information for the list of people consulted in our reports</i>	

Interviewee background

Interviewee's relevance to the SEA-PLM

Main topic

Use topic headings, not necessarily in order discussed

Subtopic

Main topic

Subtopic

Recommended follow-up

People to consult

Recommended documents/data



SEA-PLM Literature Review

May 2020

Cambridge Education
22 Station Road
Cambridge CB1 2JD
United Kingdom

T +44 (0)1223 463500
camb-ed.com

Click or tap here to enter
text.

SEA-PLM Literature Review

May 2020

Issue and Revision Record

Revision	Date	Originator	Checker	Approver	Description
1	13/05/2020	Stuart Johnson/ Attiq Sadiq	Dr Rachel Outhred	Emma Mba	Literature Review

Document reference:

Information class: Standard

This document is issued for the party which commissioned it and for specific purposes connected with the above-captioned project only. It should not be relied upon by any other party or used for any other purpose.

We accept no responsibility for the consequences of this document being relied upon by any other party, or being used for any other purpose, or containing any error or omission which is due to an error or omission in data supplied to us by other parties.

This document contains confidential information and proprietary intellectual property. It should not be shown to other parties without consent from us and from the party which commissioned it.

Contents

1	Introduction	1
1.1	The purpose and origin of this literature review	1
1.2	The evaluation of SEA-PLM for which this review is undertaken	1
1.3	Contribution of the literature review towards the evaluation of SEA-PLM	1
2	Methodology	3
3	Findings	4
3.1	Introduction to international and regional learning assessments	4
3.2	Evidence of the impact of the international and regional learning assessments	6
3.2.1	PISA	6
3.2.2	PISA Shock	6
3.2.3	Wider PISA Impact	7
3.2.4	Criticisms of PISA	8
3.2.5	SACMEQ	9
3.2.6	PASEC	13
3.3	Enablers and barriers to achieving impact	15
3.3.1	The government and the external, institutional environment in the member country	15
3.3.2	The skills and characteristics of the institutions and individuals who lead the assessments	16
3.3.3	Factors integral to the assessment, including the rules and commitments associated with participating	16
3.3.4	The nature and analysis of the learning assessment findings	17
3.3.5	The approach to disseminating the findings	18
3.3.6	Responding to and acting on the findings	19
4	Summary of findings	20
5	Bibliography	23

1 Introduction

The implementation of large-scale assessments (LSAs) has been on the rise in many countries, with around 70% of the world undertaking some mode of national or international assessment (Lietz and Tobin 2016). One of the main goals of conducting an LSA, is to determine how the education system is performing locally, nationally and internationally. In conducting such assessments, policy makers are able to check the health of an education system and determine whether reforms are working to improve learning outcomes or whether they need to pivot in a new direction.

The proliferation of learning assessments, both national and international, has led to an interest in the extent to which LSAs are having an impact on education policy and more importantly whether they are improving learning outcomes (Breakspear 2012, Baird et al 2011). This brief literature review examines the aims and objectives of selected international assessments, considers how these assessments are valued by different stakeholders, and whether the results of these assessments influence government officials to enact policy changes or wide-scale reforms.

1.1 The purpose and origin of this literature review

This literature review has been undertaken specifically as part of the process of preparing the inception report of an evaluation of the Southeast Asia Primary Learning Metrics (SEA-PLM) programme. The evaluation is commissioned by UNICEF East Asia and Pacific Regional Office (EAPRO), under contract number 43292532.

The purpose of this literature review is to provide a wider context of regional and global large-scale learning assessments within which to situate the SEA-PLM. This review is focused to that end. It draws on existing reviews and descriptions of selected key global and regional LSAs.

1.2 The evaluation of SEA-PLM for which this review is undertaken

To date, SEA-PLM has been developed and piloted in six countries, with a further five countries as observers. The first round of SEA-PLM is due to report in 2020. UNICEF and its major direct partner in supporting the survey, the Southeast Asia Ministers of Education Organization (SEAMEO), wish to take stock at this point, to inform the way forward in the next phase of SEA-PLM development and towards establishment of the survey as a long-term programme, with a periodic cycle of implementation and reporting.

This evaluation is therefore being undertaken to provide UNICEF, SEAMEO member countries and other stakeholders information to support their decision-making and planning regarding future engagement and adjustments to the SEA-PLM programme.

1.3 Contribution of the literature review towards the evaluation of SEA-PLM

The process of evaluating SEA-PLM will involve examination of the implicit and explicit evidence of rationales and expectations of stakeholders related to SEA-PLM and to the changes it is envisaged SEA-PLM may generate, and the development of the programme to date. The evaluation will also refer to the Southeast Asia regional developmental and educational context more generally.

By providing a broader frame of reference against which to set that regional level investigation and the more detailed consideration of the SEA-PLM programme, the present literature review of selected LSAs supports the development and testing of a theory of change of SEA-PLM as part of the evaluation process.

The literature review is structured in four sections. The first section provides an introduction to the literature review and its purpose, section two discusses the methodology, section three presents the findings including the impact of LSAs and the enablers and barriers for achieving impact, and section four provides the conclusion.

2 Methodology

A keyword search was conducted in the University of Oxford's SOLO library resource. Search terms included "learning assessment", "regional assessment", "PISA", "TIMMS", "PIRLS", "PASEC" and "SACMEQ". Items returned in the search that focused on technical guidelines or procedures were excluded, as were any items that focused on reporting results.

A search of grey literature was also undertaken online using google. Search terms "international assessment impact", "regional assessment impact" and the same exclusion criteria as outlined above was applied to the searches.

The authors reviewed the questions outlined in the ToR to be covered in the literature review and identified three major areas to be covered in the literature review:

- Introduction to international and regional assessments;
- Evidence on the impact of international and regional assessments; and
- Enablers and barriers to achieving impact.

The literature was reviewed, and the authors made judgements concerning the applicability of the content to speak to the above three areas. Applicable evidence was categorised according to the three major areas to be covered by the review and summaries by category.

3 Findings

3.1 Introduction to international and regional learning assessments

Since 1960, international and regional learning assessments have enabled countries to compare educational attainment across countries. However, more recently the majority of developing countries have started to conduct their own national assessments to help generate localised evidence. It is estimated that large scale assessments are being conducted in around 70% of countries, which includes a mixture of both national, regional and international assessments.

International and regional learning assessments have a variety of purposes. These include comparing levels of national achievement between countries, identifying the major determinants of national achievement and examining to what extent they are the same or differ across countries, as well as identifying factors that affect differences between countries (Howie and Plomp, 2005). Plomp (1998) suggests six different purposes for international and regional learning assessments:

- Description (or mirror) – descriptive comparisons with other countries may serve to identify particular aspects of a national system that could be considered problematic due to the extent to which they are out of line with what is found in other countries (e.g., the content of the curriculum, achievement levels of students).
- Benchmarking – setting a standard against which policymakers, the media or the general population may judge their education systems
- Monitoring – regular assessments (in order to provide trend data) of particular aspects or levels in the educational system with the purpose of making informed decisions about change when and where it is needed
- Enlightenment – a diffusion of ideas or understanding into the sphere of organizational decision-making
- Understanding – the findings can contribute to the understanding of differences between or within educational systems, which can be helpful in making decisions about the organization of schooling, the deployment of resources, and the practice of teaching.
- Cross-national research – international comparative studies may be taken as a starting point for research (“the world as a laboratory”) leading to a better understanding of the factors that contribute to the effectiveness of education.

Particular international studies may have additional purposes. For example, one of the goals of SACMEQ is to build capacity within ministries of education in order to evaluate and monitor the quality of the education being offered (Murimba, 2005). Regional studies, like SACMEQ and PASEC, are important as they compare and contrast countries which share similar economic and social conditions (Best et al 2013). While other studies, such as PISA and TIMSS, are given more prominence on the international stage and are often used by more developed nations to enact policy reforms.

Table 4.1: Summary details of the latest editions of selected international/regional learning assessments

Assessment	What measuring?	Official objectives	Geographical coverage and number of countries involved	Years of publication	Age group / Grades covered
PISA	Measures 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges (OECD 2020). Each year a different subject is measured with reading being the first subject introduced in 2010.	<ul style="list-style-type: none"> compare education system performance internationally by measuring student learning outcomes results are often used by policy makers and politicians to argue for improvement in the education system or in a specific area 	International coverage PISA I- 32 countries PISA II- 10 countries PISA III- 41 countries PISA IV- 57 countries PISA V -65 countries PISA VI - 9 countries PISA VII - 64 countries PISA VIII- 72 countries PISA IX- 80 countries PISA X- 85 countries	PISA I- 2000 PISA II- 2000+ PISA III- 2003 PISA IV- 2006 PISA V -2009 PISA VI – 2009+ PISA VII - 2012 PISA VIII- 2015 PISA IX- 2018 PISA X- 2021	15- year olds
SACMEQ	Testing Grade 6 learners and their teachers in Reading and Mathematics and the collection of contextual data, through administration of specially-designed self-completed questionnaires, on the conditions that might influence teaching and learning in schools. Since SACMEQ II, it has included linked items from other assessments including Trends in International Mathematics and Science Study (TIMSS) to allow for further analysis.	<ul style="list-style-type: none"> provide educational officials and researchers with training in the technical skills required to monitor, evaluate, and compare the general conditions of schooling and the quality of basic education generate information that can be used by decision-makers to formulate plans for improving the quality of education widely disseminate and use SACMEQ research results as the basis for policy and practice. 	SACMEQ I- 7 countries SACMEQ II- 14 countries SACMEQ III- 15 countries SACMEQ IV- 15 countries	SACMEQ I- 1999 SACMEQ II- 2004 SACMEQ III- 2011 SACMEQ IV- 2017 (ongoing)	Primary Grade 6
PASEC	Testing grades 2 and 6 in reading and mathematics at the start and end of primary education. The assessment also measures the factors that contribute to academic success.	<ul style="list-style-type: none"> The primary objective of PASEC evaluations is not the comparison of student achievement across countries, but the analysis of key factors relevant to foster educational quality attempts to derive a hierarchy of potential educational interventions in terms of their efficiency Provides technical assistance to countries to sustain national learning assessments 	PASEC I- 10 countries PASED II- 15 countries	PASEC 1- 2014 PASEC II- 2019	Primary Grades 2 and 6

3.2 Evidence of the impact of the international and regional learning assessments

A large body of evidence exists on the impact that international and regional learning assessment have on governments. Assessment programmes can help countries assess their own performance and examine whether reforms impact educational outcomes (Tamassia 2012). A number of international and regional assessments enable comparisons across countries, and each assessment has its own objectives and purpose. Some assessments predominantly focus on developed countries, however, increasingly a number of assessments have started to monitor developing countries to help enable them make more informed policy decisions. In this section, we review three well known international and regional assessments- PISA, SACMEQ and PASEC. They have all in some way impacted both developed and developing countries to reform their education policies and agendas.

3.2.1 PISA

The Programme for International Student Assessment (PISA) was launched in 2000 by the Organisation for Economic Co-operation and Development (OECD) to help countries assess whether the investments they are making in their education systems is leading to improved knowledge and skills. The OECD, has produced a 'curriculum independent' test which assesses 15-year olds on three areas reading, mathematics and science. The tests are conducted every three years, with one of the three subject areas being the focus of that year. So far, nine iterations of the assessment have taken place with the next assessment to take place in 2021.

3.2.2 PISA Shock

PISA is arguably regarded as the gold standard of educational assessments (Breakspear 2012). The assessment has received considerable attention across the globe and its results have the ability to influence senior government stakeholders to enact radical changes within their education systems. As a result, scholars (Baird et al 2011; Breakspear 2014) have coined the term 'PISA shock' which describes the response that countries have adopted as a result of the PISA rankings. Countries such as Denmark, Germany and Japan have all experienced 'PISA shock' with each country enacting large scale reforms as a result of their ranking (Breakspear 2012).

Breakspear (2014) argues that one of the reasons the 'PISA shock' syndrome is able to take root is due to the highly publicised nature of the PISA results. All major media outlets, such as the Economist, the New York Times, the Guardian etc, report on the PISA rankings and highlight the best and worst performers. This scale of media coverage can result in swaying public opinion on how the education system is performing and in turn pile pressure on governments to act quickly.

In 2000, when the first PISA assessment was launched, only three out of the 32 countries who participated implemented large-scale reforms. Germany, Denmark and Switzerland launched some form of education reforms to respond to their country's PISA ranking. In Germany, as a result of lower than expected performance, the government carried out widespread reforms which included, generating national standards, reviewing curriculum standards to include PISA competencies, and providing support for disadvantaged children especially from immigrant background (Breakspear 2012). These reforms were spearheaded after intense political debate on the 2000 results.

In Denmark, prior to implementing large scale reforms, the government first conducted an international review of the assessment to determine its credibility and simultaneously also

carried out a diagnostic on its education system to determine the gaps. After, the Danish government implemented wide-scale policy reforms which included addressing the equity divide amongst the economically disadvantaged. Additionally, the government increased the focus on national assessment and evaluation and adopted a policy of being in the top 5 PISA rankings by 2020 (Egelund 2008).

In another example Lietz & Tobin (2016) provides evidence of how Hong Kong has used the assessment results to support education reform to focus on student outcomes not only in terms of academic achievement but also in terms of non-academic outcomes such as students' social and communication skills, self-learning ability, and global outlook.

Lastly, Japan enacted wide scale reforms as a result of both its 2000 and 2003 performance. In the 2000 results Japan was regarded as a high performing country but did not perform as well as expected in reading literacy. Consequently, Japan adopted new legislation that made it mandatory for all students in lower and upper primary to participate daily morning reading sessions. In 2003, Japan's results had fallen significantly and had garnered debate on the state of its education system. Consequently, the Ministry of education abolished the low-pressure curriculum policy (Yutori) and changed its assessment practices (Tobin et al 2015, Breakspear 2012).

3.2.3 Wider PISA Impact

While the PISA results can cause external shocks to education systems, the assessment can also be leveraged as a political tool to mark a country's improvement. According to Knodel and Walkenhorst (2010), in the first two rounds of PISA the UK had appeared to perform moderately well and the government used the results to help demonstrate their performance. Little to no reforms were enacted to help improve the UK's ranking with some scholars suggesting that this was due to significant reforms being adopted only a few years earlier. Dobbins (2010) also argues that the effect of PISA results for high performing countries, such as the UK and New Zealand, is to reinforce existing policies creating the view that there is no need for a new reform agenda.

Another sign that of the impact of PISA, is the inclusion of PISA metrics in national frameworks and plans. Baird et al (2011) and Breakspear (2014) state that several countries have included PISA performance standards within their sector priorities, which includes:

- Mexico - as part of their Education sector development programme, PISA was included as a performance standard in Mathematics and reading.
- EU - included a PISA based target for low achievers within its strategic framework for European cooperation on education and training.
- UK- included the OECD targets in the importance of teaching white paper
- Japan- Used PISA competencies as well as its own national assessment framework to monitor the education system

Finally, a study conducted by the OECD asking representatives from 37 OECD countries how they use PISA results demonstrated the wide scale influence of the assessment (Breakspear 2012). The results highlighted (figure 1) that over 80% of countries have stated that their overall international rank led to changes in policy or practice. Other performance metrics, such as reading, mathematics and science also inspired countries to make improvements.

Question 18. Which aspects of your country's PISA results (across any survey round), have led to/inspired changed in policy or practice in your country?

	Yes	Partly	No
A. Overall international rank	51.4% (19)	29.7% (11)	18.9% (7)
B. Reading performance	54.1% (20)	40.5% (15)	5.4% (2)
C. Mathematics performance	43.2% (16)	43.2% (16)	13.5% (5)
D. Science Performance	32.4% (12)	54.1% (20)	13.5% (5)
E. Trend performance	40.5% (15)	29.7% (11)	29.7% (11)
F. Equity	29.7% (11)	43.2% (16)	27.0% (10)
G. Student interest, engagement, motivation and attitudes	8.1% (3)	56.8% (21)	35.1% (13)
H. Other issues (please state below)	8.1% (3)	8.1% (3)	83.8% (31)

Figure 4 PISA Impact Results

In the same survey, countries identified that they were influenced by high-performing countries and often referenced these countries in the policy making process. High performing countries, such as Finland, Australia, Korea and Canada, were cited as systems which countries looked upon to provide evidence of what works. In this instance, low performing countries become recipients of lessons which can be learnt from high performing countries. This is often referred to as policy borrowing. This practice can be useful if the contexts and education systems are similar, but caution is advised as countries may not understand the contexts in which they are borrowing policies from. Without fully comprehending the policy environment, countries can do more harm to their education systems than good.

3.2.4 Criticisms of PISA

One of the major critiques of PISA, is that it has become a high stakes assessment with significant influencing power which goes against the primary objective of assessments. The influence of PISA ranking has resulted in countries such as Australia, Denmark and Wales to adopt targets of being in the top 5 or top 20 of the PISA rankings within a defined time period. Other countries have also opted to achieve average performance by 2021 (Brazil and Thailand) (Breakspear (2014)).

A key concern shared by scholars (Breakspear 2014, Stanley 2013) is that adopting education reforms based on rank position is inappropriate as rankings can easily be influenced by small cohort changes. Stanley (2013) argues that with the introduction of PISA rankings, education has almost become similar to sporting event with nations competing to get to the top and not prioritising educational standards which should be the priority. The high stakes nature of the PISA ranking has enabled some to question whether the response to their rankings is

discouraging evidenced based decision making at a local level and that reforms are created based on achieving rank rather than the intrinsic value of teaching and learning.

Finally, as PISA only tests 15-year olds, some scholars have questioned whether a snapshot at one point in time is an appropriate measure for enacting large scale reforms. Assessments frameworks like PISA do not account for reforms which have been enacted by education sub-sectors and enacting reforms at the primary level based on PISA rankings needs to be supported by evidence which clearly identifies that the reforms are not working (Breakspear 2014). As demonstrated in the Danish case study, the PISA rankings are useful, but can only be acted upon once an internal review of the results as well as the system performance has taken place.

3.2.5 SACMEQ

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ, also referred to as SEACMEQ) was launched in 1995 and has since produced four survey iterations, the most recent taking place in 2013. The Consortium now comprises 16 ministries of education in southern and eastern Africa that work together to share experiences and expertise in developing the capacities of education planners to apply scientific methods to monitor and evaluate the conditions of schooling and the quality of education. The surveys have tested Grade 6 learners and their teachers in Reading and Mathematics and also collected contextual data on the conditions that might influence teaching and learning in schools. Since the first SACMEQ project in 1995, SACMEQ has applied virtually the same methodologies across studies and used the same instruments in order to generate valid measures of levels and changes in achievement: (a) across countries at single time points, and (b) across time points for individual countries.

Murimba (2005) summarises the impacts reported by different SACMEQ countries under five headings:

- monitoring and evaluating quality
- capacity building
- enhancing the quality of statistical and nonstatistical information systems
- policymaking and systems improvement processes
- choosing pathways to the achievement of Education for All (EFA).

3.2.5.1 Monitoring and evaluating quality and improvements to information systems

According to Murimba (2005), SACMEQ's role in monitoring and evaluating the quality of education in the region is particularly valuable because there has otherwise been a lack of data on the quality of education in member countries. The data generated by SACMEQ has given countries insight into the quality of education that their systems offer from an input, process and outcome perspective. The linkages within the different data sets collected by SACMEQ allow for comparisons against country-specific norms or expectations, across countries and over time. Ministries can therefore make assessments of their systems' performance (in terms of learning outcomes) against other countries, and against standards that they have independently set for themselves.

As well as collecting learning outcomes in the form of reading and mathematics assessments, the SACMEQ surveys have collected a larger set of input and process data about the education systems in order to provide answers to the concerns SACMEQ policymakers have expressed over the quality of their education systems (Murimba 2005). Ministries can therefore use the data to establish the extent to which the provision of different learning resources has reached

benchmarks set by policy (input variables), and can, for example, tell the frequency with which pupils are given homework, get help at home with such homework, or take extra-tuition (process variables).

Below are some examples of the influence of SACMEQ on the quality of information systems and the ability of member countries to monitor the quality of their education systems:

- Before Namibia's participation in the SACMEQ studies, education authorities relied on anecdotal evidence from education officials and so were not fully aware of the magnitude of the gaps and needs in the quality of primary education nationally and across regions (Makuwa and Maarse 2013). By participating in the SACMEQ studies, Namibia was able to make a scientific assessment of trends over time in the reading and mathematics achievement levels of Grade 6 learners for the 13 regions and at the national level. Namibia was also enabled to make valid comparisons of the performance of its own education system with the performance of other similar education systems in the southern and eastern African region. The SACMEQ I national report was widely disseminated in Namibia and was used as input in the Presidential Commission on Education, Culture and Training for its recommendations (Howie and Plomp 2005).
- In Kenya, Howie and Plomp (2005) identify that SACMEQ I led to the development of a new awareness in the Ministry of Education of the need to make major improvements in the systemic collection of data.
- In Uganda, HEART (2016) reports that some officials at the MoESTS described how the SACMEQ results are used to benchmark standards and how they learn from best practice in other countries.

As a rule, SACMEQ studies only draw samples when it has been established that the relevant education system's database is up-to-date and accurate. Where this condition is not met, time and effort are invested in the preparation of an accurate statistical database. The fact that data cleaning required such an intensive process may have implications for the quality of other data collected by governments. Murimba (2005) believes that the intensive processes and the skills and practices learned fostered a culture of quality with implications for the quality of other data collected by governments. The first component of education ministries' data to benefit from this was the education management information systems (EMIS) data. There are many concrete examples of countries that have benefited in this way.

- In Zimbabwe, where SACMEQ first began, the Ministry of Education took action following SACMEQ I to develop a comprehensive information management system.
- Botswana took advantage of the opportunities provided by SACMEQ to build its capacity for the production of an accurate and up-to-date EMIS and has subsequently used the data as a tool for diagnosing the education system's performance and for planning the development of the system.
- In Kenya, the high standards of data collection and management adopted by SACMEQ have prompted the Ministry of Education to adopt similarly rigorous data collection and cleaning practices as part of regular, established practice.
- In Swaziland, SACMEQ strengthened the EMIS and the Ministry's research function benefitting subsequent research on issues of concern to the Ministry.

3.2.5.2 Policy making and systems improvement processes

This information has equipped countries with information on the areas where learners and teachers are experiencing problems, has served as an impetus for a variety of school improvement programmes and has shaped policies that focus on quality enhancement

(Murimba 2005). The following are some of the many examples of the influence SACMEQ has made on policy decisions:

- In Kenya, Zanzibar and Zambia, lower than expected results led to the formation of national policy reviews or presidential commissions (HEART 2016).
- In Kenya, the 2002 education sector analysis made extensive use of the findings of SACMEQ I for the development of proposals aimed at addressing deficiencies observed in its education system (Murimba 2005). The findings of SACMEQ influenced the Kenyan Education Sector Support Programme, introduced in 2003. In particular, this programme allocated significant funding to the building of classrooms, providing instructional materials, and to developing non-formal education programmes for those dropping out of school. These actions emphasise the role of SACMEQ in galvanising actions in participating countries, although the precise role of new assessment data alongside previously emerging country and donor priorities is difficult to isolate (HEART 2016).
- Also in Kenya, the SACMEQ I results uncovered that a large proportion of primary school children were over-aged, which was linked with the observation that nearly two-thirds had repeated a grade at least once. The high repetition rates were in part a function of a shortage of school places and, in turn, they exacerbated the pressure for school places. It was also shown that over-age pupils tend to leave school prematurely. In response, the government put in place alternative educational opportunities for over-age pupils and it led to consideration of policy measures for dealing with grade repetition (Murimba 2005).
- In Malawi, SACMEQ I revealed that there was under-provision of educational inputs (Murimba 2005). The results from SACMEQ I were used in the national Education Policy Investment Framework to generate policy suggestions concerning the provision of teaching materials and classroom furniture for the primary education system (Howie and Plomp 2005).
- In Mauritius the SACMEQ I national report was used as a resource document in the review process of the Education Master Plan (Howie and Plomp 2005). When the prevalence of ability streaming and the scale of private tuition was highlighted, it sparked off heated national debates, and ultimately decisions were taken to regulate both (Murimba 2005).
- Mozambique's Education Sector Strategic Plan highlighted that the primary school system faced the multiple challenge of low access rates, perceived low quality, high costs of educational provision and the challenge of sustainability. It therefore used the preliminary results from SACMEQ to inform decisions on the best strategies for tackling these challenges (Murimba 2005).
- In Uganda, a gender unit was set up in the MoESTS to address weaknesses identified by the SACMEQ report, and the findings have guided the teacher recruitment process (HEART 2016).
- In Zanzibar, SACMEQ I resulted in the acceptance within the Ministry of the need to define and publish standards for the educational environment (Howie and Plomp 2005). In particular, it led to discussion on the issue of class size which, according to SACMEQ I findings, stood at 53 pupils instead of the official figure of 45. School sizes were reviewed, with excessively large schools split into smaller ones (Murimba 2005).
- In Zambia, the results of SACMEQ were used as justification and in the design of major educational programmes and new policy initiatives. They were also used in the debate and search for new interventions and monitoring mechanisms such as the National Assessment Project (Howie and Plomp 2005).
- In Zimbabwe, the results of SACMEQ I led to a number of major policy and programme initiatives, such as the Schools Rehabilitation Project (repairing classrooms, teachers houses

and constructing toilet facilities). The findings were also used in the Presidential Commission of Enquiry into Education and Training.

- In Namibia a policy was developed to share good practice and resources between clusters of schools (Tobin et al 2015, reported in HEART 2016).
- In Namibia, Makuwa and Maarse (2013) identify six ways in which the results of SACMEQ studies have been used to improve learning outcomes:
 - based on evidence from SACMEQ I and public discontent about the poor quality of education, a Presidential Commission of Inquiry on Education and Training was appointed to review the education system
 - the results of both SACMEQ I and II projects, as well as tangible actions needed in response, were shared by the SACMEQ national research team in seminars in all 13 regions. This resulted in general agreement that concerted efforts by all educational stakeholders were necessary to make these improvements.
 - as a result of the SACMEQ I and II projects, the Permanent Secretary of the Ministry of Education requested regional education authorities to develop ‘Action Plans’ aimed at addressing the low levels of reading and mathematics skills among learners. This was followed by the development of the Ministry of Basic Education, Sport and Culture’s Strategic Plan for 2001-2006, whose overarching priority was to improve equity and the quality of education.
 - educational authorities at the national, regional, inspection circuit, school cluster, school and community levels were mobilised for action.
 - the creation in 2001 of 13 education regions from the previous seven regions through decentralisation of the Ministry of Education brought education administration closer to schools
 - SACMEQ II indicated that some learners from poor communities did not have regular meals. As a result, the Ministry of Education extended the school feeding programme to more schools in poor communities to improve learners’ nutritional status and school attendance.

3.2.5.3 Capacity building

The impact of SACMEQ has also been in terms of building capacity within ministries of education to evaluate and monitor the quality of the education being offered. The training provided focuses on the skills that planning officers need in order to generate valid, evidence-based information required for policy development (Murimba 2005): research activities, such as planning and designing a research study, instrument development, test development, sampling, data collection, data capture and cleaning, data analysis, data interpretation and policy report preparation), and also other elements of the policy research process, such as the management of the consultative process that facilitates dialogue with policymakers at every stage of the research cycle, the dissemination of research findings to the different stakeholders and the creation of linkages with other partners who might benefit from the use of the information generated.

According to Murimba (2005), every SACMEQ ministry reported positively on the benefits of capacity-building. In Botswana, for example, five educational planners and statisticians were comprehensively trained in large-scale surveys. In Lesotho, SACMEQ training was linked to the Annual Training Programme of the International Institute for Educational Planning (IIEP) and has so created a stable team of highly skilled officers. In Malawi, when they failed to collect data of an acceptable quality, a team of trainers from fellow SACMEQ countries, the SACMEQ Co-ordinating Centre and the IIEP provided training in sampling, data collection and data analysis

in June 2002. By September, Malawi had put its sampling frame in order and had completed SACMEQ II data collection to required standards.

SACMEQ has also been instrumental in enhancing other elements of ministries' capacity. There is evidence from Tanzania, for example, that the hardware and software demanded by SACMEQ were put to use for other tasks, such as the capture and processing of census data and specific research studies on education.

3.2.6 PASEC

The Programme for Analysis of Educational Systems (PASEC) was launched in 1990 following the education for all conference in Jomtien, Thailand and was the first French international education assessment. Since its first data collection in 1991, assessments have been undertaken in over 20 Francophone countries not only in Africa but also in Southeast Asia (Cambodia, Laos, and Vietnam). In the early stages of the assessment, individual panel studies were carried out using different approaches which did not allow comparability. However, since 1995 national PASEC teams have been set up to ensure a robust and comparable methodology. The assessment methodology collects student data on mathematics and reading at primary grades 2 and 6 to coincide with the start and end of primary school (Hounkpodote 2018).

One of the key attributes of the PASEC assessment, is that unlike other assessments discussed in this review, the PASEC assessment objective is not to compare student achievement across countries but to understand the key factors which underscore and influence education quality (Bernard 2006). Best et al (2013) highlights that this focus of PASEC is important for policymaking as PASEC prioritises regional policy concerns in the design of the assessment which make it more likely that results would be used in policymaking.

Bernard (2006) states that the impact of the PASEC results on policy outcomes vary based on the stability of the PASEC team, the political climate and stability of the country and the ability to mobilise financial resources for the policy intervention. Bernard (2006) argues that the likelihood of PASEC results having significant leverage over governments is dependent on the results feeding directly into national education or poverty strategies and that this will in turn attract donor funding.

Impact of PASEC results

Bernard (2006) provides the following examples of the impact and influence which PASEC has had on countries.

Senegal

In Senegal one of the main objectives of the assessment was to gain insights into the effects of grade repetition on learning. Grade repetition at the time was 20% in francophone sub-Saharan Africa compared to 10% in anglophone parts of the continent. Given that PASEC is primarily administered in francophone countries, the assessment results would be of use to other neighbouring countries.

The results from the Senegalese assessment showed that a student who repeats their grade does not do significantly better when factors such as family background, school environment etc are considered. The results encouraged rigorous policy debate, aided by the PASEC team, and led to acceptance of the results by the Minister and the Cabinet. Consequently, a new policy was introduced in 2003 which prohibited repetition in grades 1-6. Despite the introduction of the ban, key stakeholders such as parents, teachers, inspectors etc still felt that repetition was the only way to help poor performing students. The teacher survey which was conducted alongside

the student assessment evidenced that 77% of teachers still believed that repetition is the best way to accelerate learning. To address this misinformation several stakeholders, including the media, publicised the results in press articles in local languages as well as French. Additionally, the Ministry of Education also introduced regional seminars to help disseminate the results. These seminars would target key stakeholders such as inspectors, teachers etc to ensure they understood the reason for the new policy.

Similar assessments were conducted in Cote d'Ivoire and Burkina Faso with results being almost identical to Senegal. However, the response to the results varied due to political instability in Cote d'Ivoire and a lack of dissemination of the findings in Burkina Faso due to lack of funding.

Guinea and Togo

In Guinea and Togo, PASEC assessments were carried out with a thematic focus on teacher training and non-civil servant teacher contracts. These issues were prioritised by governments because of prior reforms in teacher employment policies and the need to understand the impact of the policy on education quality. Both Guinea and Togo had previously enacted policies to help reduce costs in order to meet the demand for new teachers. This included hiring contract teachers, often inexperienced and paid considerably lower salaries, and only providing up to a year of teacher training in Guinea and none in Togo. These policies were enacted to help increase the teaching workforce, however little regard was paid to the impact it would have on education quality. It was anticipated that the policies would have significant impact on the education quality delivered and that an assessment was needed to determine the impact.

The PASEC results showed that new teachers, who were employed on a contract, were not performing any worse than their colleagues and, in some contexts, performed better. In line with PASEC's objective to understand the factors which influence such policies, the PASEC assessment found that the reduced time given to on-boarding contract teachers meant that the teacher training programmes were more focused on relevant content, such as practical teaching practice. Other factors such as teachers' educational attainment and teachers' desire to retain the temporary contract also led to teachers not performing worse.

The findings of the assessment were pleasing to governments because they were able to justify their reform efforts in a fiscally challenging environment. However, there was cause for concern as the findings were not able to determine why educational attainment was sub-optimal and they were of even greater concern for teachers, both contract and permanent, who aired their concerns relating to working conditions. Upon disseminating the findings, the teacher unions in both Togo and Guinea failed to attend dissemination events, however teachers, inspectors and other stakeholders agreed on the need for high-quality training for teachers to improve educational attainment. In Togo, PASEC results were distributed on a flyer by the Ministry of Education which emphasised the need for lower entry requirements for teachers, but a mandatory teacher training for newly recruited teachers. In Guinea, the results meant the continuation of training programme which was initiated under a World Bank project.

The Senegalese, Togo and Guinea experiences demonstrate that PASEC results are highly relevant for policy and decision making. Bernard (2006) argues that the relevance of PASEC is rooted in the design of the assessment which accounts for localised/regional policy concerns and that findings feed into national and education and poverty strategies. He states that a genuine interest from governments in the thematic area of the assessment can help spur on reforms in the future.

3.3 Enablers and barriers to achieving impact

The previous discussion detailed the many ways in which international and regional learning assessments have made an impact. The literature also provides many examples and lessons on the factors, decisions and contexts that can enable or hinder the achievement of that impact. Below we divide these lessons into six categories:

- The government and the external, institutional environment in the member country
- The skills and characteristics of the institutions and individuals who lead the assessments
- Factors integral to the assessment, including the rules and commitments associated with participating
- The nature and analysis of the learning assessment findings
- The approach to disseminating the findings
- Responding to and acting on the findings

It is intended that these enablers and barriers will assist policy makers, planners, assessment experts to understand the factors which enable or constrain the implementation of assessment frameworks.

3.3.1 The government and the external, institutional environment in the member country

Enablers

- The organisational culture of the host ministry, in particular if it is open to learning and adaptation, and political stability are paramount in ensuring that the assessment results are taken seriously and are the basis for policy making.
- Positioning planning functions and processes centrally within ministries and close to decision-making groups tends to result in greater impacts of assessments on policy.
- Securing strategic engagement within government – seeking champions and focusing on analysing data in areas of interest to policymakers.
- Importance of inter-linking activities with the activities of other international organizations and bilateral donors.
- The prominence of assessments can cause external shocks which provides a window of opportunity for education actors to enact large scale reforms.
- Setting targets in sector development plans and poverty alleviation strategies may support reform efforts by clarifying measurable targets for improvement and thereby mobilising system actors and shared goals.
- Learning from more successful, geographical contextual countries can help propel reform efforts.
- Consistent participation in international assessments can help policy makers understand multiple factors which affect learning.
- Wider educational reforms or political changes can create fertile ground and opportunities for using assessment results to shape new policies.

Barriers

- Enacting reforms as a result of ranking can cause detrimental effects to the education system.
- Adopting policies from other countries without understanding the context within which the policy operates can cause more harm than good.

- Need to extend involvement beyond ministry staff and academia – to include media, religious organisations, private sector, NGOs, parliaments, teachers' unions, parent groups and schools.
- Time and resource constraints can prevent governments taking part in design, data collection exercise which can cause issues in the reliability of the data or lead governments to question the methodological approaches and the findings that are produced.

3.3.2 The skills and characteristics of the institutions and individuals who lead the assessments

Enablers

- The technical, advocacy and leadership skills of the national research coordinators and other leading individuals in the national agency responsible for the assessment are critical.
- The involvement of the best experts in all elements of policy research, evaluation and assessment to provide technical leadership, particularly in the early phase, is important for setting and maintaining high standards.
- Assessment bodies to have an in-country presence to ensure buy-in of assessment results
- Important for national ministries and in-country national agencies to analyse and disseminate the results together.
- Where national research coordinators have regular access to ministries and regularly inform decision makers of the progress of assessments and their outcomes, this creates broader interest and engagement.

Barriers

- There is often a lack of communication between the Ministry of Education and the national assessment agencies in country which makes it difficult to disseminate results
- Individuals and agencies specialised in learning assessments and research may not have competencies in areas such as advocacy, social mobilisation and capacity development needed to maximise the opportunities for educational reform.
- Because of the specialised training they receive through participation in international and regional assessments, there is frequent turnover or loss of leading individuals, such as national research coordinators

3.3.3 Factors integral to the assessment, including the rules and commitments associated with participating

Enablers

- The involvement of international experts and the high technical standards and specifications give policymakers confidence in the data and gives it authority (over national data)
- The purpose of the assessment needs absolute clarity and must be realistic. This needs to drive the frequency and content. For example, an assessment designed to track national standards over time is unlikely to be suited to providing diagnostic information for teachers.
- Where countries can only withdraw data for technical reasons and before the results are known, this avoids dissemination being subject to national and political judgements.

- Making the assessment contextually appropriate for individual countries' policy contexts and ensuring that underlying factors are also studied to help policy makers make the right choices
- Working to ensure that international and regional assessments complement rather than duplicate the work of national assessments.
- Providing a high degree of confidence that the analysis process will keep to the timetable is important for persuading countries to make the investments needed to participate.
- The organising international-level institutions of the assessments should provide transparency through accessible documentation and strong communications, particularly concerning overall purposes, goals, methodologies, clear expectations and detailed timelines. This would lead to higher levels of coherence and greater benefits for participating countries.
- Ensuring students are motivated to take the assessments is crucial as it could otherwise mislead the findings

Barriers

- Participation in international and regional learning assessments involves large commitments in terms of human, material and financial resources, particularly for developing countries.
- Recent programs require an active contribution to all phases of the program—a task that for some countries is demanding due to a lack of expertise and resources.
- Many countries have a choice of assessments – national, citizen-led or regional/international (e.g. SACMEQ and PISA for Development).
- The large number of existing or planned international learning assessments is forcing policymakers to evaluate the general characteristics of potential programs in terms of scope, methodology, cost and interpretability of results.
- Regional and international assessments are less well suited to tracking national standards and barriers.
- International and regional assessments tend to require long time periods from planning and design to data collection and from data collection to the publication of findings.
- As more countries become involved in regional/international assessment exercises, the coordination of tasks becomes more complex and time-consuming, potentially holding up progress.

3.3.4 The nature and analysis of the learning assessment findings

Enablers

- The production of high quality results, based on rigorous methodologies and analysis, with valuable information about countries' relative performance and capable of diagnosing gaps and challenges and stimulating action
- The data needs to accurately represent what students achieve and it must permit valid comparisons between countries
- Contextual information is needed to provide a framework for interpreting findings on achievement
- Innovative analysis of results, pinpointing specific gaps in teaching and learning and identifying the underlying factors which contextualise and might explain findings

- The information needs to serve the practical needs of policy makers and address their own policy problems
- Carrying out follow-up diagnostic assessments targeted at assisting teachers and learners in areas needing improvement
- Workshops for data analysis and interpretation ensure that national data are accurately analysed and comparable with the international analyses while offering capacity building for national experts.
- Disseminating findings to influential stakeholders, such as UN bodies, media outlets, and other countries can help build the prominence of the assessment which in-turn can help influence countries
- A review of the assessment results and the education system should follow before reforms are adopted

Barriers

- Proposing reforms on the whole education sector based on one cohort finding can reverse any gains made in other cohorts
- Government staff capacity not built to analyse and assess findings can lead to results being unused in the policy making process
- Cross-national studies can easily lead to criticisms of unfair comparisons of national performance.
- Where responsibility for analysing assessment results lies with a department with wider responsibilities, such as producing the annual school census, competing priorities can limit the resources available to produce results in a timely fashion.
- International/regional assessment results that are from data collected two or more years previously can be dismissed as outdated.

3.3.5 The approach to disseminating the findings

Enablers

- Where results are produced at the international level, governments of participating countries need to be given regular updates on the content of materials to be published with opportunities given to express concerns or disagree.
- The production of a series of research briefs presented to ministers and other senior officials.
- Timeliness of results, when published and when disseminated, particularly when other data sources – national surveys – provide more regular commentary
- Wide dissemination and open discussion of results, including sufficient resources to do so.
- Countries playing a strong role in developing the strategy for disseminating results and participating in the dissemination process itself.
- The publication of national reports in local languages at the same time as the international release.
- Dissemination through multiple mediums (newspaper, radio, seminars etc)
- Ensuring that teachers and district education staff are aware of the assessment results and are provided with information in a way that is accessible and meaningful for their work.
- Innovative analysis of results, pinpointing specific gaps in teaching and learning
- Follow-up diagnostic assessments targeted at assisting teachers and learners in areas needing improvement

Barriers

- Dissemination events not attended by key stakeholders which lead them to questioning findings
- Dissemination materials not translated into local languages leading to local stakeholders not understanding the reasoning for reforms
- Slow reporting of findings

3.3.6 Responding to and acting on the findings**Enablers**

- Governments conducting their own reviews based on the findings to determine the way forward
- Understanding that findings are a snapshot of a cohort and not a representation of the whole education system
- Targeted holistic interventions initiated at system level, including the revision of teacher training, curriculum, standards etc, backed up by support services for administrators, heads of school, teachers, parents, local communities and all stakeholders.
- Sufficient resources are needed to act on the findings
- Teachers need support/training/targeted resources to help them to use the assessment results
- Policy borrowing both a positive and a hindrance
- Impact on educational policy and practice can be more immediate and direct in smaller countries – because the same researchers are involved in the studies and in curriculum development and policy making (short feedback loop) and because they otherwise lack information and expertise

Barriers

- Findings may not take into account the recent reforms which have been implemented which may lead governments to enact changes that are not needed or not enacting changes at all

4 Summary of findings

Large scale learning assessments are estimated to be conducted in around 70% of countries globally. These include national, regional and international assessments. International and regional learning assessments have many purposes, including comparing achievement between countries, identifying the determinants of national achievement and identifying factors that affect differences between countries. This literature review focused on three major international and regional learning assessments: PISA, SACMEQ and PASEC, drawing out lessons under two main headings:

- What impacts have been achieved by the learning assessments?
- What are the enablers and barriers in achieving those impacts?

Impacts achieved by the learning assessments

PISA is regarded by many as the gold standard of educational assessments. Its results, published every three years, receive considerable attention globally, including in major media outlets, and have been known to hold influence over senior government officials and lead to radical changes in education systems. For countries that have performed well, the PISA results can serve to reinforce existing policies and may be co-opted for political agendas. In countries that perform less well or worse than expected, the effect has been described as a “PISA shock” with large-scale reforms implemented as a result. One such example was Germany in the first PISA assessment, which went on to enact widespread reforms, including generating national standards, reviewing curriculum standards to include PISA competencies and providing support for disadvantaged children. The considerable influence that PISA now holds is seen in the inclusion of PISA metrics in national frameworks, plans and targets, however this has also received criticism. The introduction of PISA rankings has been likened to an international sporting event. Critics highlight that the focus on rankings has been at the expense of nuanced sub-national evidence and efforts to enhance the intrinsic value of teaching and learning. In addition, with PISA only testing 15-year olds, some scholars have questioned the appropriateness of using PISA as the basis for reforms across the entire education system.

SACMEQ is an example of a regional learning assessment which now brings together 16 ministries of education in southern and eastern Africa. The assessments test the reading and mathematics skills of Grade 6 pupils. Contextual information on the conditions that might influence teaching and learning is also collected. Combined, the assessments provide member countries with insights into the quality of education that their systems offer from an input, process and outcome perspective, allowing comparison cross-nationally and over time. Many and varied examples are provided in the literature of the influence that the SACMEQ has made since the first results were published in 1999.

For most member countries, SACMEQ has provided a significant step up in the quality of the data available for monitoring their education systems. It has also had knock-on effects for the quality of other data collected by governments, such as EMIS data. In terms of influences on policy making, the list of examples is long. It includes the formation of presidential commissions and national policy reviews as well as significant reallocations of funding for the construction of classrooms and the provision of classroom furniture and teaching materials. There are examples of the SACMEQ results uncovering practices and problems that were previously unrecognised, such as high repetition rates and the prevalence of over-age pupils who tend to leave school prematurely. SACMEQ has resulted in the establishment of new institutions, such as the Gender Unit in Uganda, and new national monitoring systems, such as the National

Assessment Project in Zambia. One further impact achieved by SACMEQ has been in terms of building the capacity within ministries of education. Staff have received training across the range of activities needed for rigorous evidence generation. They have also been instructed in the dissemination of findings and facilitating dialogue with policymakers and wider stakeholders.

PASEC is an international learning assessment that measures mathematics and reading levels among primary pupils in grades 2 and 6. Since its creation in 1991 by the Conference of Ministers of Education of Francophone Countries (CONFEMEN), assessments have been conducted in over 20 countries in Africa and Southeast Asia. Unlike PISA and SACMEQ, PASEC's goal is not to provide a comparison of student attainment levels across countries; rather its purpose is to understand the factors which underpin and influence education quality.

In Senegal, for example, one of the objectives of the assessment was to gain insights into the effects of grade repetition on learning. It generated evidence that grade repetition is not associated with significantly better learning outcomes, contrary to the view held by three quarters of teachers. Based on the PASEC findings, a new policy was introduced prohibiting grade repetition in primary schools, while messaging in the media and seminars with teachers and school inspectors were implemented to explain the new policy. Interestingly, similar evidence collected in other countries in the region was less effective in influencing policy, due to political instability in Cote d'Ivoire and insufficient dissemination activities in Burkina Faso.

In Guinea and Togo, the PASEC assessment was used to explore whether the policy to recruit teachers with minimal or no formal teacher training was negatively affecting the quality of education delivered in schools. Controversially, the assessments found that these new teachers were not reducing the education quality – a finding which was welcomed by the governments, but not by teacher unions. The PASEC findings revealed that the effect of the reduced teacher training was countered by a higher average educational attainment among the new teachers and also their motivation to retain their contract. It also showed that, while the new teachers had received less training, that training had benefitted from being focused on practical teaching practice. In Guinea, the results led to the continuation of the teacher training programme with funding from the World Bank. In Togo, lower entry requirements were set for teachers, coupled with mandatory training.

Enablers and barriers

There are many examples and lessons to be learned from the international and regional learning assessments and the different the ways in which they have been implemented both internationally and within individual countries. The review presents these as lists of enablers and barriers in order to assist policy makers, planners, assessment experts engaged in the planning and implementation of assessment frameworks.

At the macro-level, securing the strategic engagement of government and of key stakeholders within government is critical for success. Political instability can result in lost opportunities, however political change or wider educational reforms can create fertile soil for shaping new policies. If the host ministry has a culture of learning and if planning functions and processes centrally, experience shows that assessment results will have greater policy impacts. Extending involvement beyond ministry staff and academia, to include media, religious organisations, private sector, NGOs, parliaments, teachers' unions, parent groups and schools is also important.

The technical, advocacy and leadership skills and energies of national research coordinators and other leading individuals in the national agency responsible for the assessment are critical. These individuals can set and maintain high technical standards and help to ensure the buy-in of results. However, individuals and agencies specialised in learning assessments and research

may lack other skills, such as advocacy, social mobilisation and capacity development, that are needed to maximise the opportunities for educational reform. To maximise impact, partnerships can be formed with organisations with these strengths.

The involvement of international experts and the high technical standards and specifications of international and regional learning assessments give policymakers confidence in the data and findings produced. To increase impact, the assessment needs to be contextually appropriate for individual countries' policy contexts and underlying factors studied to help policy makers make the right choices. Participation in international and regional learning assessments involves large commitments in terms of human, material and financial resources, particularly for developing countries. Many countries now have a choice of assessments – national, citizen-led or regional/international, including a choice of different international assessments. Regional and international assessments are less well suited to tracking national standards and barriers than alternative options and they tend to require long time periods from planning and design to the publication of findings.

To be of use and achieve impact, the data needs to accurately represent what students achieve, it must permit valid comparisons between countries and contextual information is needed in order to interpret the findings. The information must serve the practical needs of policy makers and address their own policy problems. To serve teachers, follow-up diagnostic assessments can be carried out to pinpoint specific gaps in teaching and learning. There are also dangers inherent to international and regional learning assessments. Cross-national studies can be criticised for unfair comparisons of national performance. Assessments tend to focus on one or two cohorts, but they are often used as the basis for reforms across the whole education sector. Also, it can take two or more years for assessment results to be published, so that they may be outdated and not reflect more recent policies.

The wide dissemination and open discussion of results is critical for maximising the impact of assessments. Allocating sufficient resources to this stage is important. Best results require that countries play a strong role in developing the strategy for dissemination. Ministers and senior officials require specific briefings. Dissemination is best through multiple mediums with translations into national languages and different resources and approaches targeted at different stakeholder groups. If key stakeholders do not attend dissemination events or if materials cannot be read because of language barriers, findings will be questioned and there will be a gap in understanding the reasons for any reforms enacted.

There are many examples of assessment results leading to 'policy borrowing' between countries. Good practice is for governments to conduct their own reviews based on the findings to determine the way forward. Teachers also will need support and targeted resources in order to help them to use and apply the assessment results. As with the dissemination phase, sufficient resources will be needed in order to act on the findings, particularly in developing countries, and bilateral and multilateral donors have a role to play here.

5 Bibliography

- Baird, Jo-Anne, et al., *Policy Effects of PISA*, Oxford University Centre for Educational Assessment, Oxford, May 2011
- Bernard, Jean-Marc., *Managing the Impact of PASEC Projects in Francophone Sub-Saharan Africa*, CONFEMEN secretariat, Hamburg, 2006
- Best, Maura, et al., *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries*, Institute of Education, University of London, London, August 2013
- Breakspear, Simon., *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance*, OECD, Paris, Feb 2012
- Breakspear, Simon., *How does PISA shape education policy making? Why how we measure learning determines what counts in education*, Centre for Strategic Education, Melbourne, Nov 2014
- Dobbins, Macmillan., *Education Policy in New Zealand – Successfully Navigating the International Market for Education*. In K Martens, N. Alexander Kenneth, M. Windzio, & A. Weymann (Eds.), *Transformation of Education Policy*. Basingstoke: Palgrave. 2010
- Egelund, Niels., *The value of international comparative studies of achievement - a Danish perspective*. *Assessment in Education: Principles, Policy & Practice*15(3), 2008
- IIEP Learning Portal, 'Preparing the 2nd PASEC international assessment cycle', IIEP, Paris, 2018 < <https://learningportal.iiep.unesco.org/en/blog/preparing-the-2nd-pasec-international-assessment-cycle>>, accessed 11th May 2020
- HEART., *The impact of assessment results on education policy and practice in East Africa*, DFID think piece, London, 2016
- Hopkins, Pennock, Ritzen, Ahtaridou and Zimmer., *External Evaluation of the Policy Impact of PISA*, OECD, The Hague, 2008
- Howie, Sarah. and Plomp, Tjeerd., *International comparative studies of education and large-scale change*, chapter in *International Handbook of Educational Policy*, 2005, p75-99.
- Knodel, P., & Walkenhorst, H., *What's England got to do with it? British Underestimation of International Initiatives in Education Policy.*, *Transformation of Education Policy*. Basingstoke, 2010
- Lietz, P. & Tobin, M. *The impact of large-scale assessments in education on education policy: evidence from around the world*, *Research Papers in Education*, 2016
- Loizillon, A., & Uys, P., *EAPRO Meta-analysis on what works to improve learning outcomes in primary education in East Asia and the Pacific. Literature Review: Learning Outcomes in East Asia and the Pacific (Volume 1)*, 2020
- Makuwa, Demus, & Maarse, Jan., *The Impact of Large-Scale International Assessments: a case study of how the Ministry of education in Namibia used SACMEQ assessments to improve learning outcomes*, *Comparative and International education*, 2013

Murimba, Saul,. *Evaluating students' achievements: the impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)*, Prospects, 2005, p91-108.

Stanley, G,. *Reforming educational assessment: Imperatives, principles, and challenges*, Australian council for Education Research (ACER), Melbourne, 2012

Tamassia, C.V. and Adams, R.J,. *International assessments and indicators: how will assessments and performance indicators improve educational policies and practices in a globalized society?*, in The SAGE International Handbook of Educational Evaluation, California, 2012

Tobin, M., et al,. *Using large-scale assessments of students' learning to inform education policy Insights from the Asia-Pacific region*, ACER, Bangkok, 2015

