



# **Proposal to Develop Global Learning Metrics**

## **How to measure SDG 4.1**

**DRAFT**

**14 March 2016**

---

## Contents

---

Contents .....	2
List of Figures.....	5
List of Tables.....	6
I. Executive Summary .....	8
II. Introduction.....	10
2.1 Current assessment landscape .....	10
2.2 Issues in existing assessments for comparison .....	11
III. The Goals.....	11
3.1. Guiding Principles to a Global Reporting Metric .....	13
IV. The solution to assessment comparability.....	14
4.1. Linking: the technical tool to make test comparable .....	14
4.2. How does linking work? .....	15
V. The options for building a global reporting metrics.....	15
VI. Options to Global Reporting Metric.....	16
VII. Exploring options I: New test as the benchmark case.....	21
VIII. Exploring options II: Linking cross-national assessments (CNAs) .....	22
8.1. Forward-linking CNAs .....	24
8.2. Extension to link national assessments.....	25
IX. Exploring options III: Take on and enhance existing Cross-National Assessments.....	26
X. Summarizing the options .....	30
10.1. Timing .....	31
10.2. Sustainability and coverage.....	31
10.3. Soundness of methodology.....	32
10.4. Some additional thoughts .....	32
XI. The political Dimension.....	34

XII. Governance .....	37
XIII. Budget.....	37
XIV. Gateways towards 2030: from exploration to sustainability.....	38
14.1. Pre-requisites: .....	38
14.2. Possible Steps: .....	38
14.2.1. Political .....	38
14.2.2. Technical.....	38
14.2.3. Advocacy/ Consensus Building .....	38
14.3. Deliverables: What to expect in the next 2 to 3 years .....	39
14.4. Prioritization: which cycle do we start? .....	39
XV. Financial implications for 2016/2030 scenario.....	40
XVI. Capacity Building implications for 2016/2030 scenario .....	42
XVII. References .....	44
XVIII. Appendices .....	45
Appendix A.1. Cross- national assessments statistics.....	45
Appendix A.2. New learning assessment initiatives .....	49
A.2.1 LaNA .....	49
A.2.2 PfD .....	49
A.2.3 SEAPLM.....	49
Appendix A.3 Why assessments are not comparable?.....	51
A.3.1. Content framework .....	51
A.3.2. Type of items and assessment format .....	52
A.3.3. Target population .....	52
A.3.4. Data modelling and reporting .....	53
A.3.5. Contextual information .....	53
A.3.6. Technology in assessment.....	54

Appendix B. Linking Options Advantages and Disadvantages .....	56
Appendix C. Learning Outcomes Advisory Board .....	57
Appendix D. UIS Reporting Scale Initiative Steering Committee .....	58
Appendix E. Activities and Timeline .....	59
Appendix F. Detailed Staff Costs .....	65
Appendix G. Costs for the different alternatives .....	66
Appendix H – Costing Regional Modules .....	72
Appendix I. Learning Metric Partnership (LMP) and Global Alliance for Learning (GAL) .....	74
Appendix J: Notes on paired comparison .....	76

DRAFT

## List of Figures

---

Figure 1: Options to produce global measures of learning.....	16
Figure 2: Options to develop Global Learning Metric .....	17
Figure 3: Developing a new assessment .....	21
Figure 4: Possible approach to adopt a test.....	29
Figure 5. Flowchart of the process of a generic Snapshot .....	36
Figure 6. Distribution of the costs countries face for major international assessment by category	

42

DRAFT

## List of Tables

---

Table 1: Number of country participated in assessment by target.....	12
Table 2: Number of countries participating in cross-national assessments by grade and subject 18	
Table 3: Simulating Cross-National Assessments through 2030 .....	19
Table 4: Workflow for backward-linking .....	23
Table 5: Workflow for forward-linking .....	25
Table 6: Enhance an existing Cross-National Assessment.....	27
Table 7: Enhance Cross-National Assessments for global monitoring indicators .....	27
Table 8: Summary of options.....	31
Table 9: Ranking Alternatives: the Evaluation Index.....	33
Table 10: Stakeholders and their interests .....	35
Table 11: Costs of options .....	37
Table 12: Deliverables to the expected in each year under each year .....	39
Table 13: Out-of-school children and adolescents of primary and lower secondary school age, 2013 .....	40
Table 14: Indicative costs for each assessment 2016-2030 (in USD) (cost is for each 5-year period) 41	
Table A-1: Number of countries that have national assessments, by region .....	45
Table A-2: Number of countries participating in international assessments, by region.....	46
Table A-3: Regional assessments by educational level .....	47
Table A-4: Citizen-led basic learning assessments .....	48
Table A-5: Definition of literacy in various cross national assessments .....	51
Table A-6: Definition of mathematics in various cross national assessments .....	51
Table A-7: Item development.....	52
Table A-8: Last grade primary education and lower secondary education by region .....	52
Table A-9: Data Modelling and Reporting .....	53

Table A-10: Similarity and difference across two major CNAs and areas that need harmonization	55
Table B-1: Advantage(s) and disadvantage(s) of all three options.....	56
Table E-1: Activities for Alternative 1. Linking across assessments .....	59
Table E-2 : Activities for Alternative 2. Adopting an existing robust assessment .....	61
Table E-3 : Activities for Alternative 3. Creating a new assessment .....	63

DRAFT

## I. Executive Summary

---

The focus on quality of education has led to an emphasis on the measurement of learning outcomes at all levels of education. Beyond international monitoring, there is the need to support countries' statistical capacity in generating knowledge on the learning levels of their population, and develop an innovative funding strategy to raise fund for these works.

As an UN statistical agency focus on comparable data and evidence-based decision making, UIS is raising awareness of the low levels of learning and providing evidence for developing targeted approaches towards better data collection, and defining robust indicators to report progress in learning outcomes towards the 2030 goals. This UIS Global Learning Metric (GLM) initiative aims at improving learning outcomes worldwide through better use of learning assessment results for policy development. However, due to the lack of common framework and comparable data it has been difficult to have data that could be used as baseline for monitoring.

Within a very short time frame, the international community must have the data needed to monitor learning outcomes. In response, the UIS is working with technical and regional partners, to develop effective mechanism and methodological approach to produce Global Learning Metric for varied targets so countries could use to monitor and benchmark their progress towards Education 2030. This paper provides a summary of the existing assessment landscape, the issues of incomparable data and the options to attain comparable data for global monitoring concentrating on Target 4.1. It evaluates the technical options and costs in terms of time and financial resources in hope to move forward quickly in close consultation with stakeholders.

The paper presents 5 options, except for the new assessment, the other four options are potential to create indicator for the measurement points. From the considerations of cost and implementation the forward linking via Cross-National Assessments and the enhancing the Cross-National Assessments are two viable and pragmatic options. The 5 options:

1. **New test as Reference benchmark:** develop a new assessment specific to a given target population, limited to a given number of domains, and to be implemented in all countries. This option is extreme but offers a wider perspective to evaluate the other options.
2. **Backward-linking cross-national assessments (CNAs):** connect results from current assessments using the existing frameworks and reporting metrics trying to find common denominators. This implies using items and responses from past assessment to build the common denominator.
3. **Forward-linking CNAs:** link future cycles of CNAs redesigning the assessment in a way that makes them comparable. This means agreeing on a core



common framework that later on will allow to link assessments and have new reporting metrics.

4. **Forward-linking national assessments:** this is a natural extension of the forward-linking of CNAs to the national levels. Given the more than 200 countries in a world seems to be a longer terms process.
5. **Enhance an assessment:** expand or enhance the framework of an existing assessment at the cross national level taking advantage of coverage in order to include countries or regions currently not covered. This means defining a common core for the assessment that is sub set of the current test and countries today excluded (that might mean more field work).

Given the timeline and cost consideration, enhancing an assessment ranks first. The second option, i.e. forward linking using cross-national assessments may be a good alternative. The details of each options and implications are discussed in the body of the paper.

In summary, the current agenda focuses on learning and the way to get learning data through assessments require defined procedures apply to existing national and cross-national assessments. The issues that prevent comparable data: The lack of common framework and the different approaches that national and cross-national assessments take, require specific technical considerations. The paper proposes methods to link across assessments to produce pragmatic reporting metric, provides the projected cost, advantages and disadvantages of various methods, and recommendations. With the chosen option comes with needed coordination role to deal with the political landscape, financial implication and capacity building in countries.

It is hope that this paper will open up discussion among stakeholders and provide starter topics for discussion and actions.

## II. Introduction

---

The international education community has pledged to “ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes”, as part of Target 4.1 of the new Sustainable Development Goal for Education (SDG4). Yet today, it is impossible to monitor progress towards this target and the larger goal.

No single organization can produce all of the data needed to monitor SDG4. At the global level and thematic levels, the UIS has to lead the application of common methodologies and to convene the development of new ones where they do not currently exist to allow creation of comparable data, and so responding the needs established by the monitoring frameworks that are being defined at those levels, with priority focus on the SDG on education on which UNESCO is called to perform a leading role.

The first priority will be to develop the measurement frameworks for SDG4, building on the progress made thus far to define indicators, assess data availability, coverage and evaluating the existing methodologies. In addition and as agreed by the IAEG-SDG, these reliable measures are needed at each level to generate data that are comparable across time and disaggregated by age, sex, disability, socioeconomic status, geographical location (urban/rural areas) and other relevant factors<sup>1</sup>.

### ***2.1 Current assessment landscape***

According to UIS estimates, 80% of countries have conducted a national learning assessment or participated in a cross-national initiative in the last five years. Further information and statistics of the number of countries per type of assessment are provided in Appendix A.1.

In general, large-scale assessment can be divided into two categories: school-based or household-based surveys. School-based assessments include three types:

- National assessments designed to measure specific learning outcomes at a particular age or grade that are considered relevant for national policymakers.
- Cross National Initiatives (either regional or international) administered in a number of countries, based on common agreed framework, following similar procedures yielding comparable data on learning outcomes.
- Public examinations intended to certify specific learning outcomes linked to curricula and often used to select students for continuing education programs or attainment of a certain cycle

---

<sup>1</sup> See: [http://www.un.org/ga/search/view\\_doc.asp?symbol=A/C.3/69/L.9/Rev.1](http://www.un.org/ga/search/view_doc.asp?symbol=A/C.3/69/L.9/Rev.1)

On the other hand, household-based learning assessments can be used to target populations that may or may not be enrolled in or attend school. They include citizen-led assessments and any household surveys that include an assessment component in their data collection. However, in order to have data that are comparable, the existing household assessment survey may have to go through transformation to harmonize its results to the existing school based assessment survey. In the short term, the target will be mainly on students in school, i.e. the school-based assessment, and the international education community may have to take a gradual approach to assess out-of-school children.

Currently there are some cross-national assessments that are under development:

- a. LaNA (IEA’s Literacy and Numeracy Assessment)
- b. PfD (OECD’s PISA for Development)
- c. SEA-PLM (SEAMEO’s Southeast Asia Primary Learning Metric)

Except for SEA-PLM, which is a new regional initiative, LaNA and PfD are building on current initiatives, IEA’s TIMSS and PIRLS and OECD’s PISA respectively. More details of the various new initiatives are presented in Appendix A.2.

### ***2.2 Issues in existing assessments for comparison***

There is currently no framework to reconcile the differences between the various types of assessment and produce cross-nationally comparable data. In short, it is currently impossible to gain a global perspective of what children are learning because:

- Not all regions conduct assessments.
- No single measure at any education level (i.e. the last grade of primary and lower secondary varies across countries). National assessment, if exist, are given to different grade/s. Not all countries finish the same ISCED level at identical grade
- Quality and scope of national assessments vary. Without accepted global standards, national assessments are developed or implemented based on their own standards.
- Assessments are based on different methodological framework that are difficult to link and may not yield comparable results. As national assessments assess different content and use different methodology and model to report on scores it may be hard to link across assessments.

This information gap jeopardizes the global monitoring of the new global education goal and targets.

## **III. The Goals**

---

Under the SDG Goal 4, “Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all”.

For each target there are different reporting metrics according to UIS metadata collection. For example, Target 4.6 there is currently the Programme for the International Assessment of Adult Competencies (PIAAC) which covers 23 countries (data release in 2013), with additional 9 countries data to be released in 2016. Skills Toward Employability and Productivity (STEP) which is comparable to PIAAC is reporting skills for 13 low- and middle-income countries. Therefore, it provides a total coverage of 45 countries for Target 4.6.

On the other hand, Target 4.1 is the one that has many different data sources available in the market from national assessments, regional assessments to international assessments. Some countries, like South Africa, conducted all three assessments so there are abundant of information at school-age level. However, for Targets 4.4 and 4.7 little data is available.

**Table 1: Number of country participated in assessment by target**

	<b>Number of country participate in national/cross-national assessments</b>	<b>Type of assessment countries participated in</b>	<b>Name of cross-national assessment countries participated in recent (last 5) years</b>
Target 4.1	163	School-based	TERCE, PASEC, PILNA, SACMAQ, PIRLS, TIMSS, PISA
Target 4.2	6	Household-based, School-based	EAP-ECDS <sup>2</sup>
Target 4.4	20*	School-based	ICILS
Target 4.6	45*	Household-based	PIAAC, STEP
Target 4.7	38*	School-based	ICCS

\*Given that the national developed assessments in these areas are less known, we are only providing information on the known cross-national assessments.

Source: UIS.

The amount of information exist in the market for Target 4.1 therefore present a more complex situation for international educational community. For targets that have little information or no information a framework could be established base on experiences in other targets. Due to the complexity of Target 4.1 this proposal will focus on this target and develop the global learning metric, a metric that is used for reporting and reference to measure learning progress.

---

<sup>2</sup> EAO-ECDS stands for East Asia Pacific-Early Child Development Scale which is under development in Asia and the Pacific

### ***3.1. Guiding Principles to a Global Reporting Metric***

Given the amount of information exist in the market we will go by the following guiding principles:

- **No new test.** Build upon existing evidence in harmonizing frameworks, reported population, items writing, test development and data analytics from current cross-national assessments (CNAs).
- **Leverage on existing resources.**
- **No substitution** or crowding out of **national learning assessments.** Countries could (and should) continue developing their national assessments using their own standards and methodologies while using the Global Learning Metric to benchmark their performance globally.

In addition, in order to have quality data for monitoring SDG4 the following elements are needed to achieve it:

- A concrete plan to find cross-national tool or assessment to the subject area
- Well-define outcomes and related benchmarks (for instance, the definition of the minimum level or contents)
- Repeated measurement
- Agreement on approach for the different stakeholders
- Capacity development for action at the country level in the different phases
- Resources to achieve the goals

## **IV. The solution to assessment comparability**

---

There are several key issues in existing assessments:

- Low reliability and accuracy of learning assessment data in many low income countries.
- Need for data comparability and coherence within learning assessment surveys in a given country and across countries.
- Cost and effectiveness of learning assessment.
- Need for adequate funds.
- Accessibility and usability of the learning assessment data.

Furthermore, learning assessments could vary in way that may be subtle and at different stages of assessment. The details of this discussion can be found in Appendix A.3.

The solution require coordinated action at multiple levels (global, regional, and national), by multiple partners (statistical agencies, funding agencies, international organizations, civil society and the private sector), and over different time horizons.

### ***4.1. Linking: the technical tool to make test comparable***

In order use data for monitoring we need quality data, to have quality data and scores we need the following elements:

- Common framework for the assessment
- Quality assessment items
- Linguistic quality control
- Technical standards
- Item and Test design
- Sample design and implementation plan
- Standardized field operation
- Well developed, pre-tested instruments
- Well-handled data: item scoring, data entry & cleaning, sample weighting
- Scaling methodology
- Appropriate, sound, valid data analysis
- Linkage to policy response and policy-relevant reporting

All the above elements can provide check-points along the various stages of assessments to ensure outcomes obtained are reliable and valid.

In order to align existing quality assessments a solution is to technically link assessments. Linking is a series of activities used to analyze the results from two assessments so that comparisons can be made without changing any of the individual assessments.

In short, linking is a way to align tests according to a common metric. Taken to the global level this means that it would be possible for governments and citizens to understand,

for example, how a PIRLS score of 302 in 2006 (grade 4) relates to a SACMEQ score of 495 in 2007.

#### ***4.2. How does linking work?***

There are several ways to link assessments (see Appendix B): equating, statistical calibration, statistical projection and statistical moderation. Each method will have its own design and assumption. In large-scale survey assessment usually **statistical calibration** is used. Yet ultimately, the most appropriate and effective technical option will depend on a clear understanding of the purposes and uses of learning assessments results.

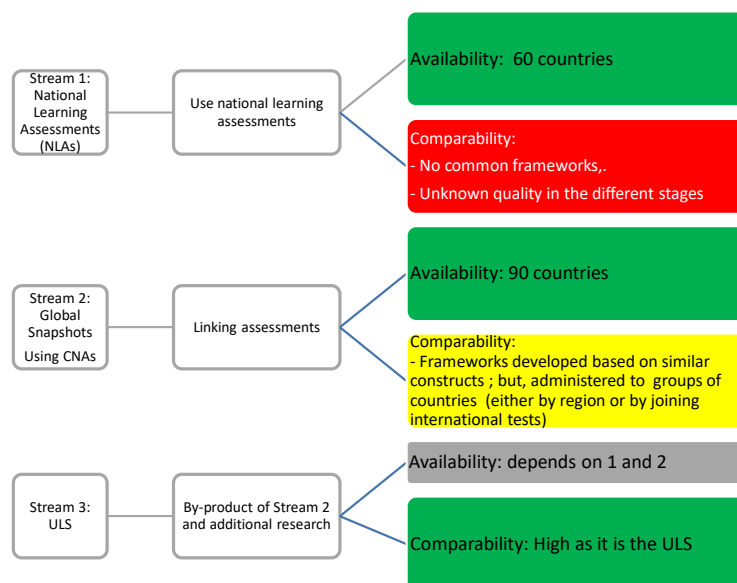
Linking can be seen as the statistical machinery that will enable countries and the larger education community to interpret evidence about student achievement. The first step is to define and agree on a common framework (Mislevy, 1992).

### **V. The options for building a global reporting metrics**

Over the past two years, the UIS has led the technical discussions on the development of a global measurement framework for learning by convening several expert groups and consultations. In particular, the UIS Reporting Scale initiatives and the UIS Learning Outcomes Advisory Board (see Appendices B and C) have helped to identify possible directions for a global strategy to measure learning outcomes. The three options are present in Figure 2.

In the following section it will develop a set of tools that link the different assessments through a robust and internationally agreed reference that would help to situate a country's progress on an **indicative pathway in the acquisition of learning outcomes**. This development work will require considerable investments in resources and time. It will also outline the different technical options to create the global learning metric while evaluating their relative strengths and weaknesses using an index to rank order each of the option.

**Figure 1: Options to produce global measures of learning**



Source: UIS.

## VI. Options to Global Reporting Metric

There are five broad options to obtain the Global Learning Metric, there are:

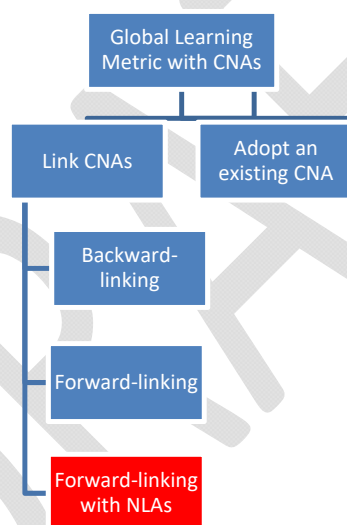
1. **New test- Reference benchmark (option 3 Universal Learning Scale as mentioned in previous section in Figure 2):** develop a new assessment specific to a given target population, limited to a given number of domains, and to be implemented in all countries. This option is extreme but offers a wider perspective to evaluate the other options.
2. **Backward-linking cross-national assessments (CNAs):** connect results from current assessments using the existing frameworks and reporting metrics trying to find common denominators. This imply, one possibility, using items and responses from past assessment to build the common denominator
3. **Forward-linking CNAs:** link future cycles of CNAs redesigning the assessment in a way that makes them comparable. This means agreeing on a core common framework that later on will allow to link assessments and produce common reporting metrics
4. **Forward-linking national assessments:** this is a natural extension of the forward-linking of CNAs to the national levels. Given the more than 200 countries in a world seems to be a longer terms process
5. **Enhance an assessment:** expand or enhance the framework of an existing assessment at the cross national level taking advantage of coverage in order



to include countries or regions currently not covered. This means defining a common core for the assessment that is sub set of the current test and developing other than countries have today excluded (that might mean more field work).

There are many options to link cross-national assessments of reading and mathematics at grades 4, 6, 8 and 10 (age 15), especially those with overlapping test frameworks. For instance, it may be possible to link two assessments (e.g. PISA and TIMSS) using a common framework with an item pool covering a wide range of tasks. Some of this exploratory work is underway. For instance, the OECD and the World Bank<sup>3</sup> have commissioned the mapping of various cross-national assessments and found many similarities. The main advantage of linking across quality assessments is the presence of established methodological frameworks.

**Figure 2: Options to develop Global Learning Metric**



Source: UIS.

There are several considerations when constructing trend for the global learning metric: align the implementation schedule; establish content framework and ensure security of items; establish reasonable minimum competencies, as defined in a learning scale, that take into account the improvement of low skills countries; and methodologies to link each snapshot across cycles.

---

<sup>3</sup> Cresswell, J., Schwantner, U., and Waters. C., (2015). Review of component skills assessed and contextual data collection used in relevant international assessments. OECD: PISA for Development Expert Paper series.

A second related issue is frequency. It is usually take time to implement policy and observe change/improvement after the first result. Therefore, normally a cycle of 3 to 5 years for the next assessment is warranted. If we have the first results or baseline at 2020 we would probably have three cycles of assessments till 2030. Three time points could probably provide educational community a trend line for each participating country.

As major cross-national assessments have different schedules it may be important to establish an agreement to align the schedule or to have the cross-national assessments to be conducting their implementation at around the same few years or adjacent years.

Methodologies will need to be identified to take into account the linking of scores across cycles. The methodologies that have been widely used by NAEP, TIMSS, PISA, etc. could be considered.

**Table 2: Number of countries participating in cross-national assessments by grade and subject**

Grade	Assessment	# of countries	Reading	Numeracy and mathematics
2	PASEC	10	X	X
3	TERCE	15	X	X
	PRILS	48	X	
4	PILNA	13	X	X
	TIMSS	63		X
5	N/A	N/A	N/A	N/A
	TERCE	15	X	X
6	PASEC	10	X	X
	PILNA	13	X	X
	SACMEQ	15	X	X
7	N/A	N/A	N/A	N/A
8	TIMSS	63		X
9 or 10	PISA (15-year olds)	65	X	X

Note: (\*) the latest cycle

Source: UIS database.

One additional piece of information is the cycle of the Cross National Assessments and the grades/ages where they evaluate. To facilitate the considerations of the proposals we have simulated the cycles of the assessment until 2030 as seen in Table 3. In some cases it has not been a systematic implementation so it is not possible to simulate.

**Table 3: Simulating Cross-National Assessments through 2030**

CNAs	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	
PISA			age 15			age 15			age 15			age 15			age 15			age 15			age 15	
PIRLS		Gr. 4					Gr. 4*					Gr. 4*					Gr. 4*					
TIMSS		Gr. 4,8				Gr. 4*,8**				Gr. 4*,8**				Gr. 4*,8**				Gr. 4*,8**				
ICILS				Gr. 8					Gr. 8													
ICCS							Gr. 8															
TERCE				Gr. 3,6																		
SACMEQ				Gr. 6																		
PASEC					Gr. 2,6				Gr. 2,6				Gr. 2,6					Gr. 2,6				Gr. 2,6
PILNA			Gr. 4,6			Gr. 4,6																

Name	Name description	Cycle
ICILS	International Computer and Information Literacy Study	Conducted first in 2013.
ICCS	International Civic and Citizenship Education Study	Conducted first in 1999 and then 2009.
PIRLS	Progress in International Reading Literacy Studies	Conducted first in 2001 then every 5-year interval, 2006, 2011.
TIMSS	Trend in International Mathematics and Science Studies	Conducted first in 1995 then every 4 years since, 1999, 2003, 2007, 2011.
PISA	Programme for International Student Assessment	Conducted first in 2000 then every 3 years since, 2003, 2006, 2009, 2012, 2015.
TERCE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE)	Conducted first in 1997 then every 7 or 9 years since , 2006 and 2013
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality	SACMEQ I in 1995 (1995-1999); SACMEQ II in 2000 (2000-2004); SACMEQ III in 2007 (2006-2011); SACMEQ IV in 2013 (2012-2014).
PASEC	CONFEMEN Programme for the Analysis of Education Systems	Although established in 1991, assessed in grades 2 and 5, from 2014 onward it will assess grades 2 and 6 at every 4 year interval. (Source. World Bank)

UNESCO Institute for Statistics –Options for Global Learning Metrics

PILNA	Pacific Island Literacy and Numeracy Assessment	A re-run of PILNA in response to low level performance in 2012 and the 2015 data will be used to confirm Literacy and Numeracy baseline.
-------	---	--

*\* Can be administered at Grades 4-6.; \*\* can be administered at Grades 8-9; Assessments after 2015 are indicative.*

Source: UIS

DRAFT

## VII. Exploring options I: New test as the benchmark case

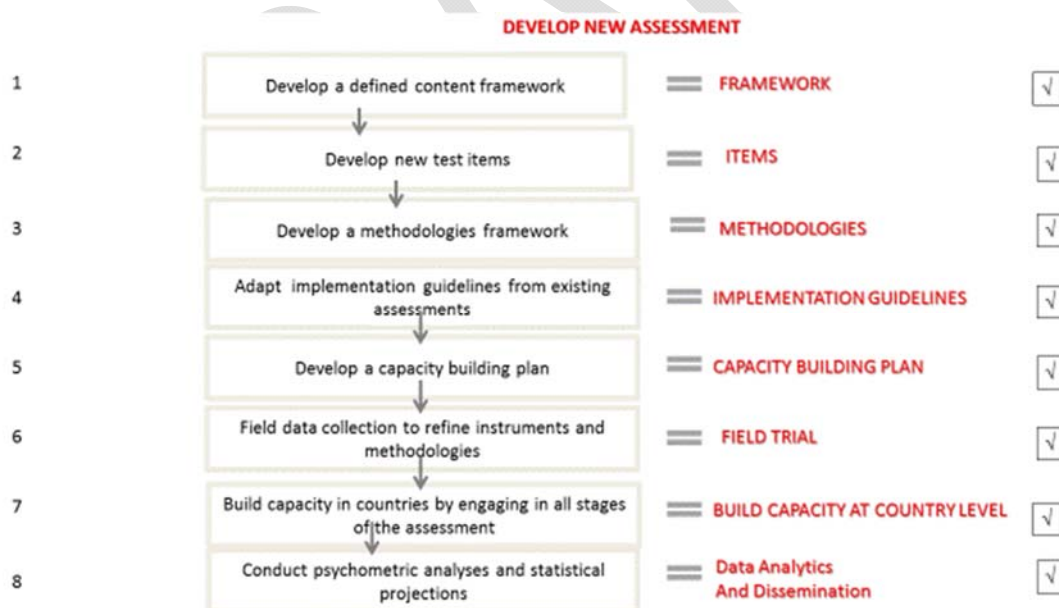
The most extreme option to produce global metrics would be the development and implementation of a new global learning assessment (specific to a given target population) in all countries.

Creating a new assessment is not the most desired or efficient option. It is by far the most expensive option and would require at least four years to implement (see Appendices E through H for estimated activities, timeline and costs).

It is certainly possible to create a new test based on experience of previous cross-national assessments while incorporating the innovative use of technology. However, there is a strong risk that countries will not be inclined to implement another learning assessment, even if the development costs were funded solely by external resources.

Why? Because implementing an assessment involves considerable financial and human resources that can even exceed the development costs. As shown in Figure 4, a new assessment would require a series of major development phases – from local capacity building to data analytics and dissemination.

**Figure 3: Developing a new assessment**



- **Costs:** Most expensive option
- **Timeline:** Minimum of five years

Option 1	Timeline (year)	
	Develop a new assessment	Field test 2018

- **Caveat and risk:** There is a risk that countries will not be inclined to implement another learning assessment, even if the development costs are funded solely by external resources. Implementation would require considerable investment at the country-level.
- **Implication for capacity building:** Since this is a new test, capacity plan could be developed for countries participate in this initiative. However the capacity plan will have to tailor to regional needs and cultural differences.
- **Immediate next step:** A platform will need to set-up to host and co-ordinate the work. This will likely be one of the activity under the Global Alliance for Learning (GAL). This will include forming a Linking Assessment Network (LAN) on all the elements mentioned in figure .

## VIII. Exploring options II: Linking cross-national assessments (CNAs)

---

There are two approaches to link CNAs: Backward-linking means connecting results from current assessments under a framework that has already been administered. Forward-linking means using a future cycle of CNAs to connect the results.

In the following sections, each approach is compared to the benchmark option, i.e. the development of a new assessment. The comparison is based on the required activities, delivery of the field trial<sup>4</sup> and main field test, as well as cost, caveats and risks, Implication for capacity building, evaluation index and immediate next step. Backward-linking

Backward-linking involves connecting results from current CNAs based on the framework used to administer the tests. Table 5 compares the activities related to backward-linking versus the development of a new test (i.e. benchmark case).

As shown in Table 5, there are considerable differences in stages 2 and 3, when significant efforts are needed to link existing procedures, frameworks, items and tools. The remaining stages are **the same** as those needed to conduct a regular assessment, except for analysis and reporting. These tasks will require a technical process to generate the common denominator needed to report the test results based on a common scale.

Clearly, the best way to evaluate this approach would involve a feasibility study and data collection to precisely understand the implication of each change or update in the process. For example, assessment items are pooled and assigned 'new' theoretical item

---

<sup>4</sup> Field trial is usually conducted to test if the intent conceptual design will provide the intent outcomes, i.e. theoretical idea is workable. Normally, in the area of assessment field trial is conducted in a smaller scale (small sample size) before actual data collection.

difficulty and content domain, these theoretical item difficulties and content domain might not always agree with their original assignment when these items were developed. Furthermore, if these items are kept for linking the new calibrated rank order might not always align with their theoretical rank order. Thus, if these items are used for linking they will change the estimation but if these items are not used the link across assessments might be weakened. Guidelines will need to establish and decision will need to make to what range of differences are acceptable to keep for or remove items from linking. Field trial could be conducted to examine the feasibility and identify appropriate methodologies to keep a stable link and retain a balance set of items contributed from each regional assessment.

**Table 4: Workflow for backward-linking**

	Stages	Backward linking	Benchmark
1	Framework	Define and agree on a content framework	Develop a defined content framework
2	Items	Obtain common item pool by pulling from all quality assessments and map items on agreed content framework	Develop new test items
3	Methodology	Harmonize a framework of methodologies from quality assessments and develop new ones	Develop a methodologies framework
4	Implementation guidelines	Adapt implementation guidelines from existing assessments	
5	Capacity building	Develop a capacity-building plan	
6	Field trial	Data collection in the field to refine instruments and methodologies	
7	Capacity building	Build capacity in countries by engaging in all stages of the assessment	
8	Data analytics	Build capacity in countries by engaging in all stages of the assessment	

Backward linking is based on huge assumption that items developed in different cross-national assessments have similar content domains (e.g. Number Operations in Mathematics) are interchangeable and comparable.

There are no common items across these cross-national assessments so field data collection is needed to collect information for statistical adjustment. One possibility (yet to be vetted) is to embed 'common items' (representative subset of items contributed by regional assessments) into representative countries who have participated in these cross-national assessments in the past. By using the relationship identified in the field data collection to re-scale the results of other countries, who have participated in the cross-national assessments, onto a common scale. Backward-linking will entail some additional costs in the initial phases (i.e. research, field trial, etc.) though total investments will be less than the cost of developing a new assessment.

- **Costs:** Less than developing a new test
- **Timeline:** Three years or less

Option 2	Timeline (year)	
		Field test
Backward-linking CNAs	2017	2018

- **Caveat and risk:** The main risk is stagnation, as backward-linking requires political leverage and effective negotiation to ensure collaboration amongst the different assessment bodies on the use of shared pool of items, especially during the initial stages of the process. As we are doing backward linking regional assessment bodies not only have to share items but also students responses to be used for re-scaling these will create additional resistance from regional assessment bodies.
- **Implication for capacity building:** Since this is backward linking, capacity plan would be for regional assessment bodies participate in this initiative. The capacity plan will have to tailor to regional needs and cultural differences.
- **Immediate next step:** Obtain consensus and make plan on collaborative work. Technical feasibility and field trial issues and potential estimation methods.

### **8.1. Forward-linking CNAs**

Instead of trying to link assessments using frameworks administered in the past (i.e. backward-linking), it is possible to look ahead and plan for future cycles of assessments. This is referred to as forward-linking.

- Through the use of special linking or anchorage packages, this option would build international comparability into the initial design stage.
- Some extra feasibility studies will be needed to generate the final test and the final forms for each of the learning domains.

As shown in Table 5, the main costs and activities of this option include developing and testing the new items to be incorporated in the next cycle of current assessment initiatives. Feasibility studies will be required, especially in relation to the data reporting and analytics.

This option will require close consultation and coordination among assessment **stakeholders and the wider international education community, especially on the part of donors and countries.** If structured effectively, these discussions can quickly lead to consensus on the specific characteristics of the framework and the selection of pilot countries per region.

- **Costs:** Less than developing a new test and backward-linking CNAs



**Table 5: Workflow for forward-linking**

	Stages	Forward linking	Benchmark
1	Framework	Define and agree on a content framework	Develop a defined content framework
2	Items	Obtain common item pool by pulling from all quality assessments and adapted and map items on agreed content framework. New items may be developed to cover those content areas that are not in existing regional assessment framework.	Develop new test items
3	Methodology	Harmonize a framework of methodologies from quality assessments and develop new ones	Develop a methodologies framework
4	Implementation guidelines	Adapt implementation guidelines from existing assessments	
5	Capacity building	Develop a capacity-building plan	
6	Field trial	Data collection in the field to refine instruments and methodologies	
7	Capacity building	Build capacity in countries by engaging in all stages of the assessment	
8	Data analytics	Build capacity in countries by engaging in all stages of the assessment	

- **Timeline:** Three to five years, depending on the speed of discussions. This process could begin immediately and be effective in 2018 based on the following timeline:

Option 3 Forward linking CNAs	Timeline (year)	
	Field test	Main survey
2017	2018/2020*	

\* Depends on the planned next cycle, which may be different by assessment.

- **Caveat and risk:** The main risk is stagnation because of the need for political leverage and effective negotiation to ensure collaboration amongst the different assessment bodies, especially during the initial stages of the process. Since this link is forward-looking so regional assessment bodies need not have to share all their items but relevant portion based on agreed common framework since new items are going to be developed.
- **Implication for capacity building:** Capacity plan would also include regional assessment bodies participating in this initiative. The capacity plan will have to tailor to regional needs and cultural differences.
- **Immediate next step:** consensus on the framework forward and make plans on harmonized methodological framework, develop estimation methodology and agree on new and understandable reporting indicators.

### **8.2. Extension to link national assessments**

National learning assessments may be possible to adapt the forward-linking methodology. This will require additional research given that for the years to come more national assessments will be aligned to global framework, this will mean more national

assessments could be linked to the global learning metric. Appednices E through H include activities, timelines and cost estimate for this option.

- **Caveat and risk:** The main risk is stagnation because of the need for political leverage and effective negotiation to ensure collaboration amongst countries and the different assessment bodies, especially during the initial stages of the process. Since the linking items are from all quality reional and national assessment bodies the main work is to establish the alignment of these pooled items with established common framework. The amount of work will increase as the number of countries involve increases.
- **Implication for capacity building:** Capacity plan would be for regional assessment bodies and countries participate in this initiative. The capacity plan will have to tailor to regional needs and cultural differences.
- **Immediate next step:** Obtain consensus on the framework, establish common framework, align pooled items with comon framework, make plans on harmonized methodologiical framework, develop estimation methodology and agree on new and understandable reporting indicators.

### **IX. Exploring options III: Take on and enhance existing Cross-National Assessments**

---

Another major option to develop global learning metric is to adapt an existing cross-national assessment. This could potentially result in considerable savings in terms of time and money as there is no longer a need for the initial investments required to link assessments. However, there is still issue on range of grade/age coverage within each ISCED level (end of primary or end of lower secondary) metric. As LaNA (Literacy and Numeracy assessment) and PISA for Development are including a range of grades (grades 4 to 6) for LaNA and ages (beyond age 15) for PfD (please see details of these assessments description in Appendix A.2). Similar approach could be taken as these cross-national assessments.

**Table 6: Enhance an existing Cross-National Assessment**

	Stages	Adopting an assessment	Benchmark
1	Framework	Define and agree on a content framework	Develop a defined content framework
2	Items	Adapt existing item pool and develop more if needed	Develop new test items
3	Methodology	Adapt a framework of methodologies and develop new ones where this assessment has not been tested.	Develop a methodologies framework
4	Implementation guidelines	Adapt implementation guidelines from existing assessments	
5	Capacity building	Develop a capacity-building plan	
6	Field trial	Data collection in the field to refine instruments and methodologies	
7	Capacity building	Build capacity in countries by engaging in all stages of the assessment	
8	Data analytics	Build capacity in countries by engaging in all stages of the assessment	

As previously explained, global learning metric is needed at the end of primary (according to UIS database 6<sup>th</sup> grade is the modal value) and lower secondary education. The assessments that cover these two points (or grades) are highlighted in red in Table 7.

**Table 7: Enhance Cross-National Assessments for global monitoring indicators**

Grade	Assessment	# of countries	Reading	Numeracy and mathematics
2	PASEC	10	X	X
3	TERCE	15	X	X
	PRILS	48	X	
4	PILNA	13	X	X
	TIMSS	63		X
5	N/A	N/A	N/A	N/A
	TERCE	15	X	X
6	PASEC	10	X	X
	PILNA	13	X	X
	SACMEQ	15	X	X
7	N/A	N/A	N/A	N/A
8	TIMSS	63		X
9 or 10	PISA (15-year olds)	65	X	X

For the end of primary education, there are four regional initiatives and a newly developed LaNA by IEA. However, except for IEA's LaNA, each one uses a different framework so there would be significant work and costs involve in linking and enhancing them as in Forward Linking using CNAs. Moreover, these four initiatives do not cover all regions and are conducted in different languages. Consequently, it is currently not possible to adopt a major initiative from these four regional assessments for the end of primary education. However, depending on the development of LaNA, which links to

IEA's TIMSS and PIRLS, may be a potential assessment to take on for end of primary education<sup>5</sup>.

A very different scenario emerges at the end of lower secondary education, with the next cycle of the PISA initiative (scheduled in 2018) prepared to cover more than 100 countries around the world. This assessment could be enhanced, possibly with the development of regional modules, in order to generate a global snapshot of learning at the end of lower secondary education<sup>6</sup>. PISA could be a potential assessment to adopt to get global metric for end of lower secondary.

The enhancement phase would probably require the modification of procedures as well as pilot-testing in countries that do not usually take part in the assessment. Instead of developing new test items, it may be possible to use a section of the test from the regional assessments mapped to the selected cross-national assessment (and the corresponding framework) to create the global metric. It will be essential to carefully structure consultations and negotiations with different countries to ensure that the resulting frameworks reflect their specific needs and realities and with reference assessment bodies, like OECD or IEA, that regional module will be add-on.

It is therefore advisable to consider the addition of regional modules to complement the larger initiative. The regional module that is tailored to regional needs and cultural difference is a good way to produce buy-in from countries. As many developing countries are finding these major cross-national assessments, like PISA, PIRLS and TIMSS, are too Euro-centric and their concerns and voice are not heard when participated in these international meetings and training workshop. Furthermore, adding a regional module to existing cross-national assessment allow us to build on reference assessment bodies' existing methodological framework: sampling, field administration and implementation. The added cost of field trial the added module will be contained within the field data collection that is going on anyway by the major CNAs. Please see **Appendix H** for details concerning the average cost of about USD 400,000 to develop a regional module.

In summary, the added value for this option is cost effectiveness since it is piggy-back onto existing assessment and we are using the infrastructure of the reference assessment with modification and update on the procedure. This option will also reduce

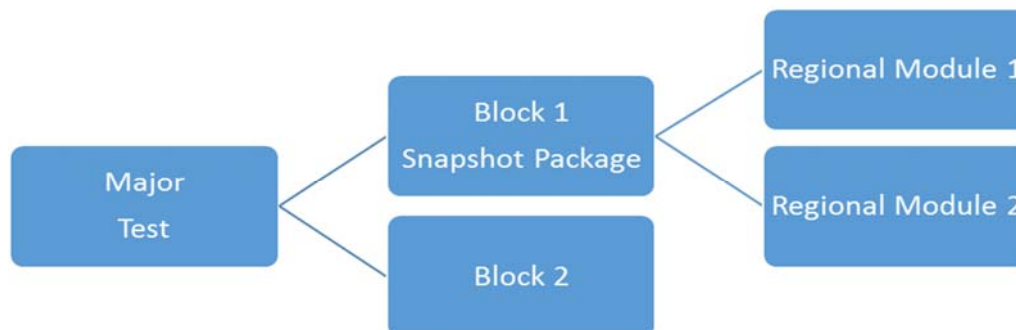
---

<sup>5</sup> Given that IEA is working on LaNA to include an easier module and is linked to current TIMSS and PIRLS. Base on LaNA proposal the initial field trial is in 2016 and main data collection in 2017, together with regular TIMSS (2015) and PIRLS (2016) data collection, IEA's TIMSS and PIRLS is positioned to be the global metric for end of primary by 2020.

<sup>6</sup> Furthermore, PISA for Development (PISA4D) is currently underway working with six developing countries: Zambia, Ecuador, Paraguay, Cambodia, Senegal and Guatemala, to understand how PISA could be implemented in developing countries. Like LaNA, PISA for Development will have an easier module (of easier items) plus some linking items from PISA to link the two assessments.

the burden of participating countries from having to administer another assessment. Figure 4 presents a possible structure to adopt a test.

**Figure 4: Possible approach to adopt a test**



- **Costs:** Less than the other options for end of primary and lower-secondary education
- **Timeline:** Three to five years depending on the next cycle of the assessment to be adapted.

Option 4 Adopting a test	Timeline (year)	
	Field test	Main survey
2016	2018/2020*	

- **Caveat and risk:** Countries that do not normally take part in the assessment may not be willing to engage or collaborate. It is important the adopted CNAs should be separated from the global metric. This option is intend to build on existing methodological framework to gain efficiency and contain cost. PISA, TIMSS and PIRLS could still do their own reporting but the global metric will have a separate and different reporting metric.
- **Implication for capacity building:** Ensure capacity building takes place at participating countries so countries will have the capacity to conduct their own assessment after engaging in these initiatives. The psychometric analyses are usually done centrally at the hired implementing organization. The education community will need to find a way to **simplify reporting** and find a way to produce results and reports that is easy to reproduce.
- **Immediate next step:** Create global metric building on the reference assessment’s framework methodology and field data collection schedule. The education community are fully in control of the added regional module and the way results are presented to policy makers: Define and set the agenda forward, consensus on common framework, and agreed on various stages of implementation including the analyses and the reporting of results and scores.

## X. Summarizing the options

Appendices E to H present detailed information about the activities, timelines and deliverables of test trials and main fields. This information is summarized in Table 8 for each of the alternative approaches.

In summary, the paper presents 5 options, except for the new assessment, the other four options are potential to create indicator for the measurement points. From the considerations of cost and implementation the forward linking via CNAs and the enhancing CNAs are two viable and pragmatic options. The 5 options:

6. **New test as Reference benchmark:** develop a new assessment specific to a given target population, limited to a given number of domains, and to be implemented in all countries. This option is extreme but offers a wider perspective to evaluate the other options.
7. **Backward-linking cross-national assessments (CNAs):** connect results from current assessments using the existing frameworks and reporting metrics trying to find common denominators. This imply using items and responses from past assessment to build the common denominator
8. **Forward-linking CNAs:** link future cycles of CNAs redesigning the assessment in a way that makes them comparable. This means agreeing on a core common framework that later on will allow to link assessments and have new reporting metrics.
9. **Forward-linking national assessments:** this is a natural extension of the forward-linking of CNAs to the national levels. Given the more than 200 countries in a world seems to be a longer terms process.
10. **Enhance an assessment:** expand or enhance the framework of an existing assessment at the cross national level taking advantage of coverage in order to include countries or regions currently not covered. This means defining a common core for the assessment that is sub set of the current test and countries today excluded (that might mean more field work).

The main advantage of forward linking is working with a divert group of assessment agencies and the essence of working on consensus. However reaching consensus has been an issue in the past if consensus does not reach it will delay the creation of initial indicators.

While enhancing CNA is efficiency gain, it does create its own risk. Many developing countries are not interested in working with existing assessment giants because of their past experience. UN agency, like the UIS, should set up a neutral platform and step-up to be the intermediary.

**Table 8: Summary of options**

Alternative	Timeline (year)		Political aspect
	Field test	Main survey	
Backward-linking	2017	2018	Agreement from international community and donors on the way forward. Identify specific characteristics (as defined by SC, TAG) and able (experiences) countries as pilot countries. Need strong and good negotiation on the collaboration
Forward-linking CNAs	2017	2018/2020	
Forward-linking NAs	2017	2018	
Adopt a test	2016	2018/2020	Use the infrastructure of the reference assessment with modification/update of procedure for countries which are not included in the group. Identify the representative countries as pilot countries for field test.
Reference benchmark – new test	2018	2019	Based on existing experience, a new assessment will not be completely new but could incorporate technology that has not been used before. Experimenting with new technology will drive up the costs for development and data collection

For us the key criteria would include: **timing** (having some data to report as quickly as possible); **sustainability going forward**; **coverage** (including the ability to include low and lower middle income countries) as well as **soundness of methodology**.

### **10.1. Timing**

To have immediate results, we could do backward linking with CNAs (option 2), find common denominator and link across cross-national assessments, like PISA and TIMSS, to gain an initial pool of over 100 countries' results. However, how best to link across PISA and TIMSS will be a discussion topic at Technical Advisory Group.

### **10.2. Sustainability and coverage**

Option 4 (forward linking with national assessment) might not be the quickest option since working with national assessments required time and resources. There is too much work need to be done to align national assessments before comparison. For example, in PISA country could be compared (via paired comparison<sup>7</sup>) only because they are on the

---

<sup>7</sup> The score metric is usually not the same across national assessments or two independent cross-national assessments, so alignment will be needed to compare the two national assessment scores on two different metrics. In other words, one score metric will need to be put onto another so that they are on the same platform. For example, if Assessment A has a score range of 0-100 with a mean of 50 and Assessment B has a score range of 200-500 with a mean of 350; score adjustment is required before comparison. Due to the range of scores the variabilities of the two assessments will be difference too. In addition, the reliability of each assessment will be difference so again the variances will need to be adjusted before it could be used in the analysis.

same scale/metric. Therefore these national assessments will need to be on the same scale/metric before any meaningful comparison could be made.

We would suggest forward linking enhancing a test (option 5) and include a regional module that could capture learners' skills of low and low-middle income countries in the region with items of relevant cultural context and use a different reporting metric, for example reports in easy-to-understand percentage (0-100) rather than any cross-national scale score (200-500). Paired comparison could be performed between countries once countries are on the same scale/metric. Additional details on paired comparison are presented in Appendix J.

### ***10.3. Soundness of methodology***

Most cross-national assessments have a defined content framework and a well-structured methodological framework compare to national assessments we should consider building on the existing available resources to create the initial learning indicators.

### ***10.4. Some additional thoughts***

The global metric should not substitute national assessments. These would point to Option (3- Forward Linking CNAs) in the medium term and Option (4- Linking NLAs) in the longer term. Option (2- Backward linking CNAs) is possibly quicker than Options (3) and (4), but also relatively expensive (and presumably the least sustainable), but may be worth taking forward in the short term if it's possible to do some backward linking relatively quickly and cheaply without field work with perhaps just a sub-set of cross national assessments. Option (5 – Enhancing an assessment) is the quickest option but not the most suitable for primary level given the large range of existing cross national assessments at primary, so might be worth considering for lower secondary in the short term.

Table 9 synthesizes discussion through the evaluation index that compares alternatives. Given the timeline and cost consideration, enhancing a test ranks first. However, there is a main concern that IEA or OECD metric will dominate the global south; especially many developing countries are not keen in joining these international assessments and have opted to participate in regional assessments. Therefore the second option, i.e. forward linking using cross-national assessments may be a good alternative. The main next step is to develop an alliance across these regional assessment bodies and a communicate strategy to inform public or participating countries of the way forward.



**Table 9: Ranking Alternatives: the Evaluation Index**

Option	Description	Cost	technical	Rank
1.Reference benchmark	Develop a new assessment specific to a given target population, limited to a given number of domains, and to be implemented in all countries. This option is extreme but offers a wider perspective to evaluate the other options	5	5	5
2.Backward-linking cross-national assessments (CNAs)	Connect results from current assessments using the existing frameworks and reporting metrics trying to find common denominators. This imply using items and responses from past assessment to build the common denominator	4	4	4
3.Forward-linking CNAs	Link future cycles of CNAs redesigning the assessment in a way that makes them comparable. This means agreeing on a core common framework that later on will allow to link assessments and reporting metrics	2	3	2
4.Forward-linking national assessments	This is a natural extension of the forward-linking of CNAs to the national levels. Given the more than 200 countries in a world seems to be a longer terms process	3	4	3
5.Adopt an assessment	Expand or adopt the framework of an existing assessment at the cross national level taking advantage of coverage in order to include countries or regions currently not covered. This means defining a common core for the assessment that is sub set of the current test and countries today excluded (that might mean more field work)	1	3	1

Source: UIS.

## XI. The political Dimension

A global metric is an invaluable instrument that will allow information to trigger learning improvement. However, like any other change of this magnitude it will mobilize stakeholders according to their interests. Empirical evidence shows that these types of changes produce conflict and there is a need to anticipate this conflicts as there will be losers and winners, and of institutions that privilege the status quo and constrain the introduction of the metrics (Grindle, 2002).

Who could be interested? We could divide the actors into several categories.

- ✓ Member States - Governmental Actors at the Central level. The principal central government actors in educational reform depends on the Prime Minister's or Cabinet chief's Office, the Presidency, the Ministries of Education and several technical institutes in each country. Indirectly, the schools for teacher training and the technical and professional schools are also involved.
- ✓ Other relevant ministries like the Ministries of Finance and Planning may be interested in the metric but would likely opt for international assessment.
- ✓ Regional Assessments centers and regional authorities bodies in general
- ✓ Donors. Foreign economic and technical assistance plays a massive role in this policy reform, both in the process which created it and in its implementation
- ✓ NGOs and other actors dedicated to education and/or social affairs and social welfare. These include political parties, professional and academic associations and unions, and student organizations. All of these actors play some role in defining public attitudes towards the formulation of education policy, although all suffer from internal organizational weaknesses as well. In our point of view the most relevant groups are the academic and professional associations and the unions.
- ✓ Teachers and Unions that feel threatened that teachers will be blamed and the comparability of results from countries in different situations.
- ✓ Parents want their children to learn and succeed.
- ✓ The Private Sector has been an actor in the development of the primary/secondary Global Metrics. Private sector include testing vendors
- ✓ Other private sector, industry in the country like the private companies in the country would value the learning level in the country where it will benefit from a productive workforce.

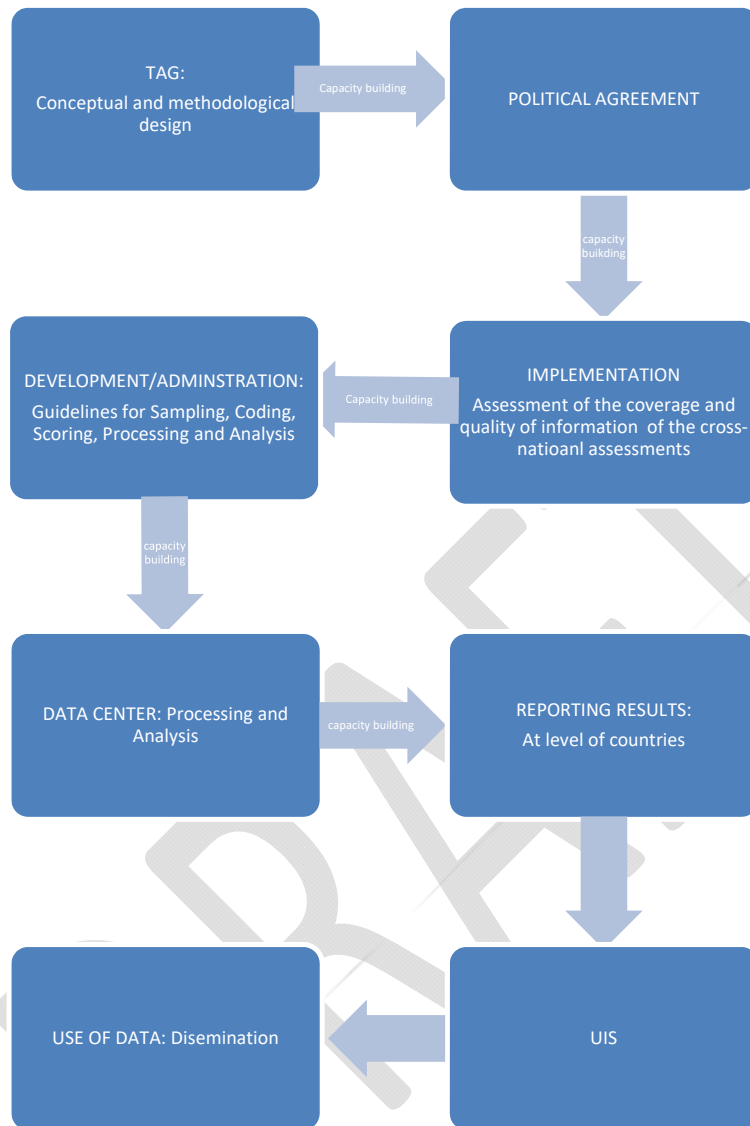
Table 10 summarizes the main interests of stakeholders and the benefits that they will gain in participating the global initiative. Many dimensions might be taking into account like greater visibility/profile; financial gain; enhanced technical capacity/skills and control.

**Table 10: Stakeholders and their interests**

Stakeholders	Stakeholders' interests and benefits
<p>Testing agencies (IEA Data processing center, Analyses and Reporting unit, Pearson, ETS, ACER)</p> <p>Regional Assessment bodies (LLECE, PASEC, SACMEQ, SEAPLM, PILNA)</p>	<p>Implementing agency of international assessments, like IEA, ETS, ACER, etc. will be the potential beneficiary in the SDG learning agenda as the international community move from collecting administrative data to learning outcomes data. Even though data collection (sampling, administration, scoring, etc.) are done at country level, the psychometric analyses (scaling and linking) need to be done at international professional agency to ensure and maintain reliable metric for reporting and trend. The final data will be housed at the UIS after the data have been processed by the selected international agency.</p> <p>Regional assessment bodies will benefit by engaging and participating in the global initiative through a gain in visibility and financial support from donors' community. The functions of the regional assessment bodies are two-folds: (1) Liaison with participating countries and help countries to build technical capacity in learning assessments in the region that are culturally appropriate. This will include relating the framework, training countries in data collection, engaging with countries in reporting and analyses for policy development. (2) Liaison with the international agency that do the psychometric analysis by ensuring a clean data set for analyses.</p>
<p>Countries</p>	<p>The countries who participate in the global initiative will gain in technical capacity on learning assessment as they will be actively involve in the adaptation of items, engage in the methodological framework of implementation (sampling, administration, scoring and data capture), and using data for analyses (for policy development) with the supports from international assessment community.</p>
<p>General public (teachers and Union, parents, private industry, NGO)</p>	<p>General public will gain by having an educated and learnt population in term of literacy, civic participation, and productivity. Except for the content framework, they will less likely be participating in the development process.</p>
<p>Donors</p>	<p>Donors will likely have a say in all processes of the work. They could use their influence to direct the outcomes. They will likely be clear on where they stand (align to their funding policy). They will be in the best position to negotiate with cross-national assessment agencies, especially those cross-national assessment agencies that have been difficult.</p>

A flow chart of the process that could be applicable almost to all cases shows the needs of agreements and clear rules in every step in the process.

**Figure 5. Flowchart of the process of a generic Snapshot**



In all cases, UIS needs to have a good communication strategy (good understanding of technical issues and able to explain in clear and simple terms) and a good working relationship with the cross-national assessment bodies. In all alternatives, the coordinating agency will need to work with all cross-national assessment bodies, though at different levels of engagement with each one.

The co-coordinating agency will mainly have to liaison with the TAG and selected CNAs assessment bodies on technical issues and work with regional assessment bodies on the adaptation, development and implementation of the adopted assessment. UIS needs to work closely with TAG and regional assessment agencies on all aspects of the development.

## XII. Governance

The initiative will be hosted by Global Alliance for Learning.<sup>8</sup> The Global Alliance of Learning includes a broad range of education stakeholders, e.g., Member States, International and regional assessment initiatives, International development agencies and bi-lateral donors, civil society organisations and other stakeholders.

## XIII. Budget

The budget has been estimated for a period of three to five years under the assumption that there is an operational team in the UIS to co-ordinate global activities.

**Table 11: Costs of options**

Alternative	Costs	Timeline (year)		Political aspect
		Field test	Main survey	
Backward-linking	\$5,951,156.00	2017	2018	Agreement from international community and donors on the way forward. Identify specific characteristics (as defined by SC, TAG) and able countries for pilot work. Need strong and careful negotiation on the collaboration
Forward-linking CNAs	\$3,951,156.00	2017	2018/2020	
Forward-linking NAs	\$4,401,156.00	2017	2018	
Adopttest	\$3,295,260.00	2016	2018/2020	Use the infrastructure of the reference assessment with modification/update on procedure for countries which are not normally in the group. Identify representative countries as pilot countries for field testing.
Reference benchmark – new test	\$8,985,260.00	2018	2019	Based on existing experience, the new assessment will not be completely new but could incorporate technology that has not been used before. Experimenting with new technology will drive up the costs for development and data collection in the field

*Note: If the process is started in 2016 by setting up the SC and TAG and with consensus on the alternative, then the implementation procedure for test data collection in the field can be harmonized and adapted. With appropriate pilot countries, the implementation procedure adapted from existing methodologies, and the successful running of field tests, then the main survey could be undertaken in the following year, i.e. 2018. This is under the assumption that there are few modifications and updates after the field test. The*

<sup>8</sup> The difference between Learning Metric Partnership (LMP) and the Global Alliance for Learning is described in Appendix I.

*main delay may be from political commitment from countries which entails finalizing MOUs and financial commitment from donors.*

Source: UIS.

## **XIV. Gateways towards 2030: from exploration to sustainability**

---

### **14.1. Pre-requisites:**

- Set-up the Global Alliance for Learning (GAL)
- Set-up the Secretariat, the Steering Committee (SC) and the Technical Advisory Group (TAG) for this particular initiative, building on existing experts from the UIS Reporting Scale Initiative and Learning Outcomes Advisory Board members.
- Secure sufficient funding and staff

### **14.2. Possible Steps:**

The way forward is presented for three areas: Political, Technical and Advocacy.

#### **14.2.1. Political**

- Mobilize stakeholders to endorse best approach for linking; and to produce a new way of reporting for international monitoring. This will include:
  - Develop a framework of collaboration with stakeholders
  - Secure Memorandum of Understanding
  - Form regional alliance and partnerships (stakeholders, assessment bodies, NGO, academia)
  - Mobilize regional resources to leverage implementation at country level

#### **14.2.2. Technical**

- Commission literature reviews and research on the design of assessments, list challenges and best way forward
- Develop methodology for linking assessments
- Establish a working research network to develop the design, and evaluate and test the methodology
- Mobilize field data collection to validate design and to test the methodology with selected countries and stakeholders
- Develop and/or update guidelines for field data collection for global implementation
- Develop guidelines for database, and tools for data analysis

#### **14.2.3. Advocacy/ Consensus Building**

- Ensure good communication
- Produce reports for policy development and advocacy
- Ensure countries' ownership and knowledge transfer

- o Set up mechanisms for the implementation, maintenance and expansion of the database based on adopted linking methodology

**14.3. Deliverables: What to expect in the next 2 to 3 years**

Table 12 summarizes the deliverables to be expected for each of the alternatives. In all cases there are two core elements that need to be delivered in 2016

1. Framework of measurement for each cycle
2. Definition of the Data Quality Assessment Frameworks (DQAF)

**Table 12: Deliverables to the expected in each year under each year**

	2016	2017	2018
Option 2: Backward linking with CNAs	<ol style="list-style-type: none"> <li>1. Consensus on the collaboration with the cross-national.</li> <li>2. Framework of measurement for each cycle.</li> <li>3. Definition of the DQAFs.</li> </ol>	<ol style="list-style-type: none"> <li>1. Field test using existing test design and administrative guidelines to see if there is context effect for a number of representative countries across the regions.</li> </ol>	<p>Possibly, no main survey is needed. Develop reporting metric if it goes well</p>
Option 3: Forward linking with CNAs	<ol style="list-style-type: none"> <li>1. Consensus on the collaboration with the cross-national.</li> <li>2. Framework of measurement for each cycle.</li> <li>3. Definition of the DQAFs.</li> <li>4. New assessment design with regional module.</li> </ol>	<ol style="list-style-type: none"> <li>1. Assemble test based on adapted test framework.</li> <li>2. Adapted administrative guidelines.</li> <li>3. Field test.</li> <li>4. Updated test and administrative guidelines.</li> </ol>	<p>Main survey and data collection. Develop reporting metric.</p>
Option 5: Adopt an assessment	<ol style="list-style-type: none"> <li>1. Framework of measurement for each cycle.</li> <li>2. Define DQAFs.</li> <li>3. Develop new assessment design with regional module.</li> </ol>	<ol style="list-style-type: none"> <li>1. Assembled test based on adopted test framework</li> <li>2. Administrative guidelines.</li> <li>3. Field test data.</li> <li>4. Updated test and administrative guidelines.</li> <li>5. Definition of the reporting metric for the regional module with the main module as the linking blocks.</li> </ol>	<p>Main survey and data collection.</p>

**14.4. Prioritization: which cycle do we start?**

In this moment the two points of measurement in the list are End of Primary and End of Lower Secondary so discussion will be restricted to this. Depending on the approach that the international education community selects, the technical process could start with either education level although from the point of view of the tests available for lower secondary seems to be the easiest level to start.

However it is not difficult to argue that starting at a lower level would be less costly and more effective to improve not only learning and reduce school drop-out. Table 13 shows that there are important reasons to start before the end of primary school.

**Table 13: Out-of-school children and adolescents of primary and lower secondary school age, 2013**

UIS region	Primary school age						Lower secondary school age					
	Number (million)			Rate (%)			Number (million)			Rate (%)		
	M	F	MF	M	F	MF	M	F	MF	M	F	MF
Arab States	2.2	2.7	4.9	10.3	13.3	11.8	1.5	2.2	3.7	13.5	20.8	17.0
Central and Eastern Europe	0.4	0.3	0.7	4.1	3.7	3.9	0.4	0.4	0.7	3.7	3.9	3.8
Central Asia	0.2	0.2	0.4	6.1	6.7	6.3	0.2	0.2	0.5	5.9	7.5	6.7
East Asia and the Pacific	3.5	3.1	6.6	4.2	4.1	4.1	4.3	3.4	7.7	9.0	7.8	8.4
Latin America and the Caribbean	2.1	2.0	4.1	6.6	6.4	6.5	1.3	1.2	2.5	7.0	6.5	6.7
North America and Western Europe	1.1	1.1	2.2	4.4	4.3	4.3	0.4	0.3	0.8	2.7	2.3	2.5
South and West Asia	5.5	4.9	10.3	6.0	5.8	5.9	13.7	12.6	26.2	25.6	25.7	25.6
Sub-Saharan Africa	13.4	16.7	30.1	18.1	22.9	20.5	10.9	11.9	22.7	32.6	36.4	34.5
World	28.4	30.9	59.3	8.3	9.7	9.0	32.6	32.2	64.9	16.9	17.8	17.3

Source: UIS Data Centre, December 2015

If the community selects option 5 - adopting an assessment - one could start with PISA/TIMSS at the end of lower secondary since it is easier to negotiate and obtain a Memorandum of Understanding with one assessment body then with several cross-national assessment bodies and as long as it is possible to successfully advocate to the countries that the CNAs used as a base is not the test but the new module adapted to regional and local characteristics is, this is the case if, for instance, PISA or TIMSS are used. Communication is key in making clear they are only as platform for collecting relevant data.

However if the international community selects option 3 - forward linking - the most reasonable and sustainable option so far, based on cross-national assessments it would be optimal to start with the end of primary education since the current cross-national assessments are mainly collecting data at the end of primary education. Economies of scale could be obtained as there are already developed test frameworks, constructs and methodologies to work with.

In any case, it will vital to ensure a structured initial meeting with the Steering Committee and Technical Advisory Group members to hash out these initial issues and make the necessary decisions.

## **XV. Financial implications for 2016/2030 scenario**

The initial investment to generate and adapt the methodological framework (e.g., content, items, guidelines, administration, tools, and analysis) and reporting metric in the development phase will vary depending on the approach chosen. The cost will be



substantially reduced once the methodological framework is established and operationalized.

The future costs will mainly involve data collection costing approximately USD 500,000 for data collection in field test and main survey per country and potential international costs for an implementing agency to conduct the psychometric analysis to scale the data across countries and international co-ordination cost across regions. Once countries have implemented the assessment, the implementation cost of the subsequent will likely decrease since training and capacity building have carried out in countries. The implementation costs could be reduced to about USD 300,000 per country for data collection in the main survey only.

It may be possible not to conduct a field test in the second cycle if the methodological framework and assessment design remain the same. There will still be international costs for scaling the data and international co-ordination cost; however, the implementing agency might achieve economies of scale by having a large group of countries conducting the assessment at about the same time and the regional agencies are familiar with procedures. The international costs will therefore have to be negotiated with the implementing agency.

The costs in table 14 are indicative costs for each 5 year period. Assuming the first 5 years (2016-2020) is the development phase so there is no country cost involved unless a pilot test is warranted. The second 5 years (2021-2025) is the actual implementation across approximately 150 countries on 1 education level (e.g. end of primary). Even though there are over 200 countries and territories given that some countries are in conflicts and/or not able to conduct any survey or assessment, we will remind optimistic that there will be approximately 150 countries or territories that we could get some initial results.

**Table 14: Indicative costs for each assessment 2016-2030 (in USD) (cost is for each 5-year period)**

	Development	Country's implementation	international scaling cost <sup>9</sup>	international co-ordination cost	total international cost
2016-2020	4,000,000			300,000	4,300,000
2021-2025		75,000,000	1,500,000	1,500,000	3,000,000
2026-2030		45,000,000	750,000	750,000	1,500,000

Note: (\*) total international= (development+international scaling cost + international coordination cost)

Source: UIS.

<sup>9</sup> Assuming international scaling and co-ordination cost is USD 10,000 each per country for the first cycle and USD 5,000 each per country for the second cycle.

If countries conduct assessment across two education levels, the cost might not be double since there are some cost saving in term of setting up the infrastructure, training, administration, analysis and reporting. The main cost will be the size of sample collected. This cost will vary from country-to-country due to its geography in order to have representative sample.

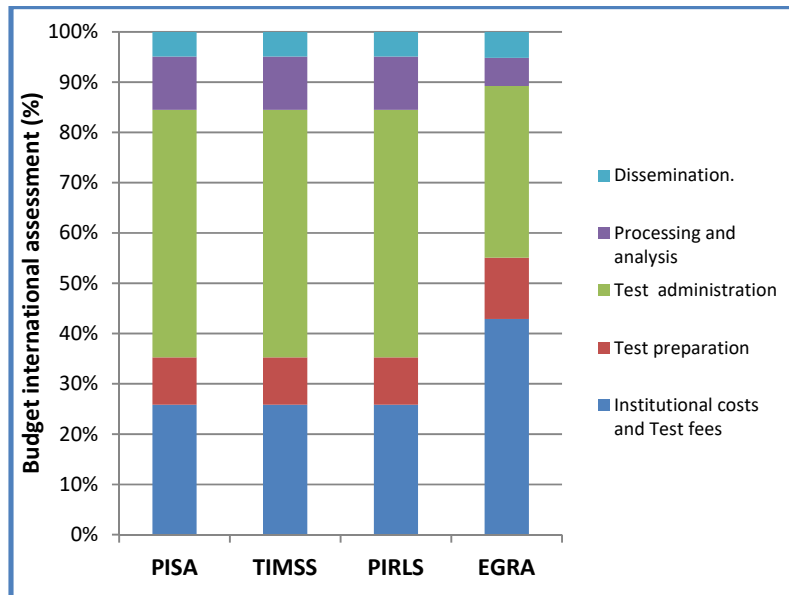
## XVI. Capacity Building implications for 2016/2030 scenario

Some of the biggest challenges developing countries face is not only financial but the constraints in terms of human resources to administer in the different stages of the implementing large scale assessments. Capacity building is key for improving the quality, timeliness and deliver by helping to improve the statistical capacity of countries and other partners.

Cross-national assessment agencies, especially regional assessment agencies, could support national institutions to build capacity in countries. Within the Global Alliance of Learning governance structure capacity building is considered a big part of the structure to help countries build capacity and align their results to the global scale for monitoring. More details of the structure and the work involved under each branch are presented in the Global Alliance for Learning concept notes.

What are the total costs and what is the structure of these costs by category **Figure 6** provides examples of international assessments and the distribution of their costs. It shows that test fees of some well-known international tests account for only one-third of the budget needed by a country. This costing (based on current observed costs) is also most likely underestimating the costs for data analytics, dissemination and use of information that are still not developed enough.

***Figure 6. Distribution of the costs countries face for major international assessment by category***



Source: UIS estimated, based on Wagner, et al. (2011). *How much is Learning Measurement Worth? Assessment Costs in Low-Income Countries*, UWEZO (2011) *Improving Learning Outcomes*.

These costs excludes the costs associated to the infrastructure of testing that in many countries is very much needed and includes from computers to dedicated officials and setting the human capital capacity. The setting up of the infrastructure in country is initially high but once the infrastructure has been set-up this fixed cost will reduce over the number of assessments conducted in the country. The variable of sample size will be the main cost over the long run if countries would like to collect more data for more detailed policy development. This cost has been estimated by Jean Marc Bernard as 5 million USD per country.

## **XVII. References**

---

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

Cresswell, J., U. Schwantner and C. Waters (2015). “Review of component skills assessed and contextual data collection used in relevant international assessments”. OECD PISA for Development Expert Paper series. Paris: OECD.

Jia, Y., G. Phillips, L.L. Wise, T. Rahman, X. Xu, C. Wiley and T.E. Diaz (2014). 2011 NAEP-TIMSS Linking Study: Technical Report on the Linking Methodologies and Their Evaluations (NCES 2014-461). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Mislevy, R. J. (1992). Linking Educational Assessments: Concepts, Issues, Methods, and Prospects. ETS Policy Information Report. Princeton: Educational Testing Service.

OECD (2015). Universal basic skills: What countries stand to gain? Paris: OECD Publishing.

DRAFT

## XVIII. Appendices

### Appendix A.1. Cross- national assessments statistics

**Table A-1: Number of countries that have national assessments, by region**

Region	Number of countries	Areas of assessment
East Asia & Pacific	22	Literacy, Reading, Language, Mathematics, Sciences, Social studies, English, Life skills, Computer studies – Information and Communications Technology (ICT), Physical education, Civics
Europe & Central Asia	41	Literacy, Writing, Reading, Language, Mathematics, Sciences, Social studies, French, Life skills, History, Computer studies ICT, Geography, Physical education.
Latin America & Caribbean	26	Literacy, Writing, Reading, Language, Mathematics, Sciences, Social studies, French, History, Computer studies ICT, Geography, Physical education., Civics
Middle East & North Africa	16	Literacy, Language, Mathematics, Sciences, Social studies, French, Life skills, Computer studies ICT
North America	3	Writing, Reading, Language, Mathematics, Sciences, Social studies, History, Computer studies ICT, Geography, Civics
South Asia	8	Literacy, Language, Mathematics, Sciences, Social studies, History, Computer studies ICT, Geography
Sub-Saharan Africa	30	Literacy, Language, Mathematics, Sciences, Social studies, French, Life skills, Computer studies ICT
<b>TOTAL</b>	<b>146</b>	-

**Table A-2: Number of countries participating in international assessments, by region**

Region	TIMSS 2010-2011	PIRLS 2010-2011	ICILS 2013	PISA 2012	CIVED 1 or 2
East Asia & Pacific	10	6	4	13	2
Europe & Central Asia	30	26	13	37	24
Latin America & Caribbean	2	3	2	8	2
Middle East & North Africa	16	9		5	1
North America	2	2	1	2	2
South Asia	-	-	-	-	
Sub-Saharan Africa	3	2	-	-	
<b>TOTAL GENERAL</b>	<b>63</b>	<b>48</b>	<b>20</b>	<b>65</b>	<b>31</b>

Source: Based on UIS Catalogue of Learning Outcomes and World Development Report.

DRAFT

**Table A-3: Regional assessments by educational level**

Organization responsible for assessment	Assessment	Number of countries	Grade	Area	Year
Conference of Education Ministers of Francophone Countries across the World (CONFEMEN)	Programme on the Analysis of Education Systems of the CONFEMEN (PASEC) VII-IX	11	Grades 2 and 5	French and mathematics	2004-2009
	PASEC 2014	10	Grades 2 and 5 or 6	French and mathematics	2014-2015
Educational Quality and Assessment Programme (EQAP)	Pacific Islands Literacy and Numeracy Assessment (PILNA)	14	Grades 4 and 6	Literacy and Numeracy	2012
	PILNA	13	Grades 4 and 6	Literacy and Numeracy	2015
Latin American Laboratory for Assessment of the Quality of Education (LLECE)	Primer Estudio Regional Comparativo y Explicativo (PERCE)	13	Grades 3 and 4	Reading and mathematics	1997
	Segundo Estudio Regional Comparativo y Explicativo (SERCE)	16	Grades 3 and 6	Reading, mathematics and science	2006
	Tercer Estudio Regional Comparativo y Explicativo (TERCE)	15	Grades 3 and 6	Reading, mathematics and science	2013
Ministries of Education of the Southern and Eastern Africa	The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) I	7	Grade 6	Reading literacy	1995-1999
	(SACMEQ) II	13	Grade 6	Reading literacy	2000-2004
	(SACMEQ) III	14	Grade 6	Reading literacy	2006-2011
	(SACMEQ) IV	14	Grade 6	Reading literacy	2012-2014

**Table A-4: Citizen-led basic learning assessments**

Programme	Age range (in years)	Languages in which children were tested	Area
Annual Status of Education Report (ASER) India	5-16	20 languages	Reading and arithmetic
ASER Pakistan	5-16	Urdu, Sindhi, Pashto, English	Reading and arithmetic
UWEZO Kenya	6-16	Kiswahili, English	Reading and arithmetic
UWEZO Tanzania	7-16	Kiswahili, English	Reading and arithmetic
UWEZO Uganda	6-16	English	Reading and arithmetic
Beekungo Mali	5-14	French, Bamanankan, Bomu, Fulfulde	Reading and arithmetic
Jangandoo Senegal	5-18	French, Wolof, Pulaar	Reading and arithmetic

Source: *Citizen-Led Basic Learning Assessments. An Innovative Approach.*  
[http://www.worldbank.org/content/dam/Worldbank/Event/education/2013\\_Citizen-led%20Learning%20Assessments%20Oct%2030.pdf](http://www.worldbank.org/content/dam/Worldbank/Event/education/2013_Citizen-led%20Learning%20Assessments%20Oct%2030.pdf)



## ***Appendix A.2. New learning assessment initiatives***

### **A.2.1 LaNA**

LaNA, initiated in 2015, is designed for countries that have not yet participated in IEA TIMSS or IEA PIRLS. The assessment is given at the end of primary education (grades 4-6 depending on the country and the countries' education system and curriculum) and would enable the countries to compare their literacy and numeracy achievement with TIMSS and PIRLS countries at fourth grade.

The newly developed IEA Literacy and Numeracy Assessment (LaNA), will link to the IEA TIMSS and IEA TIMSS Numeracy assessments conducted in 2015 and the IEA PIRLS and IEA PIRLS Literacy assessments conducted in 2016. Thus, the participating countries will obtain international comparisons of their numeracy and literacy achievement with countries who have participated in TIMSS and PIRLS from all over the world.

### **A.2.2 Pfd**

In an effort to make participation in PISA more accessible and relevant for countries – particularly developing countries - the OECD has developed the PISA for Development (Pfd) project, in consultation with development partners.

PISA for Development, initiated in 2013, is designed as a collaborative pilot project that will implement in collaboration with multiple stakeholders and participating countries. Like PISA, Pfd covers three subject areas (Reading, Mathematics and Science) and targeted at student age 15 or slightly older.

The project aims to enhance the policy relevance of PISA for developing countries through the development of enhanced PISA survey instruments and data collection methods that are more relevant for the contexts found in developing countries but which produce results on the same scale as the main PISA assessment and are therefore internationally comparable.

### **A.2.3 SEAPLM**

SEAPLM, initiated in 2013, is designed to assess learning achievement for primary school age (initially at age 10) across several content areas (reading, writing, math, global citizenship). It is hope to be used for system level monitoring on education equity and quality. It has a common tool (translated into different national languages) for each domain but at the same time allows explorations of cross-national variations in the South East Asia regional context.

It also provides tools for research and analysis which can inform policy making/policy dialogue on issues related to education reform and improvement of curriculum and its relevance to ultimately improve quality and students' learning outcomes.

It is hope that the assessment will inform discussions and plans for harmonization across ASEAN – and provide insights for developing ASEAN qualification frameworks and standards.

DRAFT

**Appendix A.3 Why assessments are not comparable?**

**A.3.1. Content framework**

Depending on the curriculum in country, national assessments tend to have different content coverage in the assessed grade. The way that even the domains are defined set the first area of difference. Tables 2 and 3 reproduce the differences in the definition of literacy and mathematics.

**Table A-5: Definition of literacy in various cross national assessments**

PISA 2000	PISA 2009/2015	PIRLS	SACMEQ	STEP
Reading literacy is understanding, using and reflecting on written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society.	Reading literacy is understanding, using, reflecting on and <b>engaging with written texts</b> , in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society.	"the ability to understand and use those written language forms required by society and/or valued by the individual. .".	the ability to understand and use those written language forms required by society and/or valued by the individual	“Understanding, evaluating, using and engaging with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential”

Source: UIS based on Cresswell, Schwantner and Watters (2015).

**Table A-6: Definition of mathematics in various cross national assessments**

PISA 2015	TIMSS	SACMEQ
Mathematical literacy is an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts.	"knowing, applying and reasoning	"The capacity to understand and apply mathematical procedures and make related judgements as an individual and as a member of the wider society".

Source: UIS based on Cresswell, Schwantner and Watters (2015).

### A.3.2. Type of items and assessment format

The assessment may use different types of assessment format; some countries use only multiple-choice items while others may use a combination of multiple-choice and constructed-response items. The process shows how the items are developed.

**Table A-7: Item development**

PISA 2015	LLECE	SACMEQ
<ul style="list-style-type: none"> <li>▪ Item generation</li> <li>▪ Panelling items</li> <li>▪ Cognitive trial .</li> <li>▪ Field trial</li> <li>▪ Main study selection</li> </ul>	<p>Uses the expert group approach in which a group of experts calls for submission of items.</p> <p>TERCE is based on a curriculum analysis, specification tables.</p> <p>Item development involves specialists from almost all countries</p>	<p>Items are developed by a panel of subject specialists drawn from all the 15 participating school systems</p>

Source: UIS based on Cresswell, Schwantner and Watters (2015).

### A.3.3. Target population

Assessment may be given to different grades, the target population may vary; some countries want to assess mid-education level, some at end of education level, while others at both mid- and end of education levels. Furthermore, the number of years of schooling (or duration of schooling) may vary from country to country.

For example, some countries have 6 years of primary so mid- and end of education level may be grades 3 and 6 respectively, while others have 4 years of primary so the mid- and end of education level may be grades 2 and 4 respectively. Table 5 reflects the case for the two points at where the global assessment should be made; in none of the ISCED 1 or ISCED 2 end of cycle ends in the same grade.

**Table A-8: Last grade primary education and lower secondary education by region**

Region	Last grade primary education					Last grade lower secondary education					Total
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	
East Asia & Pacific	1	5	27	2	-	1	7	18	8	1	35
Europe & Central Asia	20	11	18	3	1	-	11	34	7	1	53
Latin America & Caribbean	-	2	30	9	-	-	6	24	11	-	41
Middle East & North Africa	2	4	15	-	-	-	1	18	2	-	21
North America	-	-	3	1	-	-	1	2	1	-	4
South Asia	-	5	1	2	-	-	4	2	1	1	8
Sub-Saharan Africa	-	3	35	10	-	-	6	18	22	2	48
<b>Total</b>	<b>23</b>	<b>30</b>	<b>129</b>	<b>27</b>	<b>1</b>	<b>1</b>	<b>36</b>	<b>116</b>	<b>52</b>	<b>5</b>	<b>210</b>

Source: UIS.

### A.3.4. Data modelling and reporting

In term of data modelling and reporting some countries may use more sophisticated model, like the item response theory to scale and report scores, while others may use simple classical theory descriptive statistics, like the proportional correct to report scores that isolates from the different level. Given the differences the reporting scores will be difference in scale or metric.

**Table A-9: Data Modelling and Reporting**

	PISA	PIRLS/TIMSS	SACMEQ	LLECE
Cognitive reporting scale	Formerly Rasch Model (1-PL for multiple choice item and Partial Credit model for construct response item). Currently data has been rescaled to a 3-PL and Generalized Partial Credit model (GPCM).	2- or 3-PL model for multiple choice item and Generalized Partial Credit model for constructed response item	Rasch model	Rasch model
Performance/Proficiency level	6 levels of proficiency level with level 6 is the advance level and level 1 is the basic level.	4 levels of proficiency level with level 4 as the advance level and level 1 as the basic level.	8 levels of proficiency level with level 8 as the advance level and level 1 as the basic level.	5 levels of proficiency level with level 4 as the advance level and below 1 as the basic level.

Source: UIS.

### A.3.5. Contextual information

Contextual information is usually collected at the national or cross-national assessments for in-school assessments or household assessment surveys. The information could use to support policy-related analyses of the results, and inform the design and development for better in-country and regional level policy. Contextual information collect could be different from country-to-country or region-to-region but should include a few common characteristics, like gender, grade, age, location, socio-economic background and disability status, for monitoring progress.

### A.3.6. Technology in assessment

Technology in assessment has improved over the years which enable a more dynamic assessment design. With improved psychometric modelling which enables reasonable good estimation with smaller number of items and target population, it allows the adoption of different implementation platform and operational procedures.

For example, the use of computer-based adaptive test where the respondents are given a set of test items based on their existing skills so they do not have to sit through a long test. The use of computer or tablet as the presentation platform for the test takers will allow the selection of set of items to be embedded in the platform, scoring of cognitive items and coding of background information to be done automatically, and improve the efficiency of data processing for further analysis. Furthermore, technology also allows the use of authentic cognitive items, like simulation to put items into a more realistic situation.

As the technology advances different national and cross-national assessments have updated their programmes to incorporate new technology in their assessments. For example, the U.S. National Assessment and Educational Progress (NAEP), PISA, and PIAAC, to name just a few. However, not all assessments implement the advance technology. In addition, the use of technology in assessment comes with a cost in development and assessment design. Therefore it is important to set priority on what is the most needed features in assessments and budget the development cost accordingly. Are cross national-assessments comparable?

Compare to national assessments, cross-national assessments build their test based on similar framework and methodologies. The differences among cross-national assessments are less different than among national assessments. Table 7 below shows the areas across two major cross-national assessments that need harmonization.

Between PISA and PISA for Development, the instruments (test and background questionnaire) are enhanced to take into consideration of the developing country context, with easier items from PISA and fit-for-policy purpose background questions. With those common items from PISA, PISA for Development will be brought onto the same scale as PISA. With the enhanced background questions the developing countries can report on contextual issues that are closed to their interests.

**Table A-10: Similarity and difference across two major CNAs and areas that need harmonization<sup>10</sup>**

	IEA			OECD		
Stage	TIMSS	PIRLS	LaNA	PISA	PISA for Development	Area need harmonization
Target population	Grades 4 and 8	Grade 4	Grades 4-6	age 15	age 15 and slightly older	
Content coverage	Mathematics and Science	Reading	Reading and Mathematics	Reading, Mathematics and Science	Reading, Mathematics and Science	Consensus, pool and align selected content areas
Scaling methodology	IRT models with 3-PL for multiple choice items and Generalized Partial Credit model for constructed response item. Both TIMSS and Numeracy portion of LaNA, PIRLS and Literacy portion of LaNA will be on same scale.			IRT models with 3-PL for multiple choice item and Generalized Partial Credit model for constructed response item. Both PISA and PfD will be on same scale.		
Proficiency level	4 levels of proficiency level with level 4 as the advance level and level 1 as the basic level.			6 levels of proficiency level with level 6 is the advance level and level 1 is the basic level.		Need consensus to harmonize if want to link across the two

Source: UIS.

<sup>10</sup> Comparison is made between two major international assessments.

**Appendix B. Linking Options Advantages and Disadvantages**

**Table B-1: Advantage(s) and disadvantage(s) of all three options**

		Advantage(s)	Disadvantage(s)
Option 1	Statistical projection	Simple. Database exists. Quick for initial global monitoring.	Might have biased results if assumptions for statistical projection do not hold. On reference assessment’s reporting metric. Countries might not agree with projections of education data using economic data.
	Link across assessments – linking assessment to reference assessment	Own schedule. Could rely on existing framework of methodologies.	Might have biased results if framework for methodologies varies greatly between linking and reference assessment. On reference assessment’s reporting metric.
Option 3	Link to common item pool	Own schedule. New reporting metric. Could rely on existing framework for methodologies with minor adaptations.	Politically challenging. Resource-intensive if new metric is preferred but only this part needs development.



***Appendix C. Learning Outcomes Advisory Board***

Jean-Marc Bernard	Global Partnership for Education
Pierre Brochu	Council of Ministers of Education, Canada
Marguerite Clarke	World Bank
Luis Crouch	RTI International
Ariel Fiszbein	Inter-American Dialogue
Sylvie Grenier	Statistics Canada
Sheren Hamed	Jordan Education Initiative
Seamus Hegarty	University of Warwick
Anthony Nitko	University of Pittsburgh
Baela Raza Jamil	Idara-e-Taleem-o-Aagahi (ITA)
Ismael Sanz Labrador	National Institute for Educational Assessment of the Ministry of Education, Culture and Sports, Spain Universidad Rey Juan Carlos
Ralf St Clair	University of Victoria
Toziba Masalila	SACMEQ Coordinating Center
Michael Ward	Organisation for Economic Co-operation and Development

***Appendix D. UIS Reporting Scale Initiative Steering Committee***

Jim Ackers	UNICEF
Manos Antoninis	UNESCO
Ed Barnett	DFID
Penelope Bender	USAID
Jean-Marc Bernard	Global Partnership for Education
Moritz Bilagher	UNESCO LLECE
Hannah Birdsey	DFAT Australia
Michael Bruneforth	Austrian Federal Institute for Education
Marguerite Clarke	World Bank
Dirk Harstedt	International Evaluation Association
Irwin Kirsch	Education Testing Service
Sylvia Linan-Thompson	University of Texas
Jacques Malpel	PASEC
Anthony Nitko	University of Pittsburgh
James Pellegrino	University of Illinois
Abbie Raikes	UNICEF
Wilima Wadhwa	(ASER Center
Michael Ward	Organization for Economic Co-operation and Development
Kentaro Yamamoto	Education Testing Service

**Appendix E. Activities and Timeline**

**Table E-1: Activities for Alternative 1. Linking across assessments**

List of Activities	2016				2017				2018			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1. Secretariat, Steering Committee (SC) formed and Technical Advisory Group (TAG) members identified	X											
2. TAG formed and Implementing Organizations (IO) identified		X										
3. Agreement by SC on content framework or skills and competencies that children should have at the end of: primary education and lower secondary education	X											
4. SC negotiates and agrees with national and/or regional assessment body on the use of their assessment for linking		X										
5. IO to pull common item pool from all quality (national or regional) assessments and map each item to the content framework		X	X									
6. IO to harmonize a framework of methodologies that are culturally-sensitive to different regions			X	X								
7. IO to develop flexible but robust implementation guidelines			X	X								
8. IO to develop a capacity-building plan for countries participating in the linking exercise			X	X								

9. SC with the support of IO builds capacity in countries and engages countries in all stages of linking exercise					X	X	X	X	X	X		
10. Countries conduct pilot testing and IO fine-tune the framework of methodologies and conducts data collection in the field							X	X	X	X		
11. IO conducts psychometric analyses and performs statistical projection										X	X	
12. SC and countries discuss the results											X	X
13. IO conducts projection of results for countries which did not participate in the linking exercise but participated in regional assessments where items are pooled											X	X
14. SC advocates the linking exercise and conducts further discussion with international education community on the reporting and creation of indicators												X

DRAFT

**Table E-2 : Activities for Alternative 2. Adopting an existing robust assessment**

List of Activities	2016				2017				2018				2019				2020	
	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2	Q 3	Q 4	Q 1	Q 2
1. Secretariat, Steering Committee (SC) formed and Technical Advisory Group (TAG) members identified	X																	
2. TAG formed and reference assessment and/or Implementing Organizations (IO) identified. The IO could be the reference assessment body		X																
3. SC negotiates and agrees with the identified assessment body on the use if their assessment for adaptation		X																
4. IO works with the reference assessment body to adapt the existing item pool from the reference assessment and develops more items to fill the gaps where necessary		X	X															
5. IO works with the reference assessment body to adapt a set of methodological frameworks and develops others that are missing from the reference assessment to ensure cultural sensitivity for different regions			X	X														
6. IO develop flexible but robust implementation guidelines using the reference assessment implementation plan			X	X														

7. SC works with IO to develop capacity-building plan for countries that choose to participate in the adapted assessment			X	X															
8. SC with the support of IO builds capacity in countries and engages countries in all stages of the assessment					X	X	X	X	X	X									
9. Countries pilot the procedure and IO fine-tunes the framework of methodologies and conducts data collection in the field							X	X	X	X									
10. IO engages the reference assessment organization in psychometric analyses and statistical moderation										X	X								
11. SC and countries discuss the results											X	X							
12. IO works with the reference assessment body on statistical moderation of the results for countries not participation in the adapted assessment, but participating in the reference assessment												X	X						
13. SC agrees on a new reporting metric and works with IO on the creation of indicators for reporting													X	X					
14. SC advocates the methodology and further discusses the results with international community																X	X	X	

**Table E-3 : Activities for Alternative 3. Creating a new assessment**

List of Activities	2016				2017				2018				2019				2020			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1. Secretariat, Steering Committee (SC) formed and Technical Advisory Group (TAG) members identified	X																			
2. TAG formed and the Implementing Organizations (IO) identified		X																		
3. SC to agree on content framework or skills and competencies that children should achieve at the end of: primary education and lower secondary education.	X																			
4. IO develops items based on the agreed content framework		X	X																	
5. IO develop a framework of methodologies that are culturally-sensitive for different regions			X	X																
6. IO develop flexible but robust implementation guidelines for countries			X	X																
7. IO develop capacity-building plan for countries that participate in the new assessment			X	X																
8. SC with the support of IO builds capacity in countries and engages countries in all stages of the assessment					X	X	X	X	X	X	X	X								

UNESCO Institute for Statistics –Options for Global Learning Metric

<p>9. Countries conduct pilot testing and IO fine-tunes the instrument and framework of methodologies, then conduct data collection in the field</p>								X	X	X	X	X	X							
<p>10. IO engages in psychometric analyses, designs a reporting metric and generates performance description based on the new items</p>														X	X	X				
<p>11. SC and countries discuss the results</p>																X	X			
<p>12. SC advocates the new assessment and further discusses with international education community on reporting and reaction of indicators</p>																		X	X	X

DRAFT



**Appendix F. Detailed Staff Costs**

Projected budget for ULS Stream 2 staff cost												
Employees Working on Project/Program	FTE	Estimated Time Spent					Amount					Estimated Cost Total
	Standard Cost *	2016	2017	2018	2019	2020	2016	2017	2018	2019	2020	
Director (HQ) - D2	285,480	5%	5%	5%	5%	5%	14,274	14,274	14,274	14,274	14,274	71,370
Communication Officer (HQ) - P4	206,960	10%	10%	10%	10%	10%	20,696	20,696	20,696	20,696	20,696	103,480
Programme Specialist (HQ) - P3	172,380	100%	100%	100%	100%	100%	172,380	172,380	172,380	172,380	172,380	861,900
Research Assistant (HQ) - G6	59,700	100%	100%	100%	100%	100%	59,700	59,700	59,700	59,700	59,700	298,500
IT Associate Project Officer (HQ)	150,020	10%	10%	10%	10%	10%	15,002	15,002	15,002	15,002	15,002	75,010
<b>Total Salary Cost</b>	<b>N/A</b>	<b>N/A</b>	<b>N/A</b>	<b>N/A</b>	<b>N/A</b>	<b>N/A</b>	<b>282,052</b>	<b>282,052</b>	<b>282,052</b>	<b>282,052</b>	<b>282,052</b>	<b>1,410,260</b>

**Appendix G. Costs for the different alternatives**

Projected budget for ULS Stream 2 activities cost for Alternative 1 (linking assessment)					
Activity: Other Professional Services	2016	2017	2018	Total	Notes
Consultation with regional assessment bodies	35,000	35,000	35,000	105,000	Online platform and/or in person consultations
Consultations	20,000	20,000	20,000	60,000	Specific topic consultation with advisory board members
Research cost in 2016	100,000	0	0	100,000	Appointed research or commission papers, depending on the outcomes of first SC and Secretariat meeting
Field data collection development cost in 2017	0	300,000	0	300,000	International cost, guideline to harmonize procedures to reference assessment, need not develop new procedures
Field data collection country cost in 2017-2018		4,000,000		4,000,000	Country implementation cost to test framework of methodologies 1-2 countries per (5 or less) regions.
Translation cost for documents	40,000	40,000	40,000	120,000	Translating documents to 4 other languages, if necessary (French, Spanish, Arabic, Russian)
<b>Total</b>	<b>195,000</b>	<b>4,395,000</b>	<b>95,000</b>	<b>4,685,000</b>	
Activity: Travel, Conferences & Meetings	2016	2017	2018	Total	
Staff Travel	20,000	20,000	20,000	60,000	Staff travel for additional consultation; travel to SC or TAG meetings if meetings not in Montreal
Conferences	10,000	10,000	10,000	30,000	Presentation at conference to rise visibility and awareness of the project
Steering Committee (SC) meetings	20,000	20,000	20,000	60,000	Organizing of SC meetings once a year
Technical Advisory Group (TAG) meetings	40,000	40,000	40,000	120,000	Organizing of TAG meetings once a year
Implementation Training Workshops	0	100,000	50,000	150,000	3 workshops (USD 50,000 each), include regional bodies and participating countries. # workshops depend on the amount of adjustments
<b>Total</b>	<b>90,000</b>	<b>190,000</b>	<b>140,000</b>	<b>420,000</b>	
<b>Total cost across three years</b>	<b>5,105,000</b>				
<b>Estimated cost per year (spread over 3 years)</b>	<b>1,701,667</b>				

UNESCO Institute for Statistics –Options for Global Learning Metric

<b>Projected budget for ULS Stream 2 activities cost for Alternative 1.1 (forward linking assessment)//CROSS NATIONAL ASSESSMENT</b>					
<b>Activity: Other Professional Services</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>Total</b>	<b>Notes</b>
Consultation with regional assessment bodies	35,000	35,000	35,000	105,000	Online platform and/or in person consultations
Consultations	20,000	20,000	20,000	60,000	Specific topic consultation with advisory board members
Mapping items from cross-national assessments	50,000	0	0	50,000	Adapting existing items for the item pool. Assume USD 10000 per Cross National Assessment and there are 5 CNAs
Design anchor blocks to be embedded in assessments	50,000	0	0	50,000	Assemble linking items from item pool that is representative to content framework coverage, appropriate booklet design and item location.
Field data collection development cost in 2017	0	300,000	0	300,000	International cost, harmonize or adapt existing procedures, need not develop new procedures
Field data collection country cost in 2017-2018	0	2,000,000	0	2,000,000	Country implementation cost to test framework of methodologies 1-2 countries per (5 or less) regions.
Translation cost for documents	40,000	40,000	40,000	120,000	Translating documents to 4 other languages, if necessary (French, Spanish, Arabic, Russian)
<b>Total</b>	<b>195,000</b>	<b>2,395,000</b>	<b>95,000</b>	<b>2,685,000</b>	
<b>Activity: Travel, Conferences &amp; Meetings</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>Total</b>	
Staff Travel	20,000	20,000	20,000	60,000	Staff travel for additional consultation; travel to SC or TAG meetings if meetings not in Montreal
Conferences	10,000	10,000	10,000	30,000	Presentation at conference to rise visibility and awareness of the project
Steering Committee (SC) meetings	20,000	20,000	20,000	60,000	Organizing of SC meetings once a year
Technical Advisory Group (TAG) meetings	40,000	40,000	40,000	120,000	Organizing of TAG meetings once a year
Implementation Training Workshops	0	100,000	50,000	150,000	3 workshops (USD 50,000 each), include regional bodies and participating countries. # workshops depend on the amount of adjustments
<b>Total</b>	<b>90,000</b>	<b>190,000</b>	<b>140,000</b>	<b>420,000</b>	
<b>Total cost across three years</b>	<b>3,105,000</b>				
<b>Estimated cost per year (spread over 3 years)</b>	<b>1,035,000</b>				

UNESCO Institute for Statistics –Options for Global Learning Metric

<b>Projected budget for ULS Stream 2 activities cost for Alternative 1.2 (forward linking assessment)- NATIONAL ASSESSMENTS</b>					
<b>Activity: Other Professional Services</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>Total</b>	<b>Notes</b>
Consultation with regional assessment bodies	35,000	35,000	35,000	105,000	Online platform and/or in person consultations
Consultations	20,000	20,000	20,000	60,000	Specific topic consultation with advisory board members
Mapping items from national assessments	500,000	0	0	500,000	Adapting existing items for the item pool. Assume USD 5000 per country and there are 100 countries.
Design anchor blocks to be embedded in assessments	50,000	0	0	50,000	Assemble linking items from item pool that is representative to content framework coverage, appropriate booklet design and item location.
Field data collection development cost in 2017	0	300,000	0	300,000	International cost, harmonize or adapt existing procedures, need not develop new procedures
Field data collection country cost in 2017-2018	0	2,000,000	0	2,000,000	Country implementation cost to test framework of methodologies 1-2 countries per (5 or less) regions.
Translation cost for documents	40,000	40,000	40,000	120,000	Translating documents to 4 other languages, if necessary (French, Spanish, Arabic, Russian)
<b>Total</b>	<b>645,000</b>	<b>2,395,000</b>	<b>95,000</b>	<b>3,135,000</b>	
<b>Activity: Travel, Conferences &amp; Meetings</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>Total</b>	
Staff Travel	20,000	20,000	20,000	60,000	Staff travel for additional consultation; travel to SC or TAG meetings if meetings not in Montreal
Conferences	10,000	10,000	10,000	30,000	Presentation at conference to rise visibility and awareness of the project
Steering Committee (SC) meetings	20,000	20,000	20,000	60,000	Organizing of SC meetings once a year
Technical Advisory Group (TAG) meetings	40,000	40,000	40,000	120,000	Organizing of TAG meetings once a year
Implementation Training Workshops	0	100,000	50,000	150,000	3 workshops (USD 50,000 each), include regional bodies and participating countries. # workshops depend on the amount of adjustments
<b>Total</b>	<b>90,000</b>	<b>190,000</b>	<b>140,000</b>	<b>420,000</b>	
<b>Total cost across three years</b>	<b>3,555,000</b>				
<b>Estimated cost per year (spread over 3 years)</b>	<b>1,185,000</b>				

UNESCO Institute for Statistics –Options for Global Learning Metric

<b>Projected budget for ULS Stream 2 activities cost for Alternative 2 (Adapt an existing assessment)</b>							
<b>Activity: Other Professional Services</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>Total</b>	<b>Notes</b>
Consultation with regional assessment bodies	35,000	35,000	35,000	35,000	35,000	175,000	Online platform and/or in person consultations.
Consultations	20,000	20,000	20,000	20,000	20,000	100,000	Specific topic consultation with advisory board members
Research cost	0	10,000	0	0	0	10,000	Appointed research or commissioned papers
Develop field data collection procedure cost	100,000	-		0	0	100,000	International cost, assume harmonizing existing procedures, need not develop new procedures
Country field test data collection cost	350,000	350,000				700,000	Country implementation cost to test framework of methodologies 1-2 countries per (5 or less) regions. Maximum 10 countries.
Translation cost for documents	40,000	40,000	40,000	40,000	40,000	200,000	Translating documents to 4 other languages, if necessary (French, Spanish, Arabic, Russian)
<b>Total</b>	<b>545,000</b>	<b>455,000</b>	<b>95,000</b>	<b>95,000</b>	<b>95,000</b>	<b>1,285,000</b>	
<b>Activity: Travel, Conferences &amp; Meetings</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>Total</b>	
Staff Travel	20,000	20,000	20,000	20,000	20,000	100,000	Staff travel for additional consultation; travel to SC or TAG meetings if meetings not in Montreal
Conferences	10,000	10,000	10,000	10,000	10,000	50,000	Presentation at conference to rise visibility and awareness of the project
Steering Committee (SC) meetings	20,000	20,000	20,000	20,000	20,000	100,000	Organizing of SC meetings once a year
Technical Advisory Group (TAG) meetings	40,000	40,000	40,000	40,000	40,000	200,000	Organizing of TAG meetings once a year
Implementation Training Workshops	0	0	50,000	50,000	50,000	150,000	3 workshops (USD 50,000 each), include regional bodies and participating countries. # workshops depend on the feasibility studies outcomes.
<b>Total</b>	<b>90,000</b>	<b>90,000</b>	<b>140,000</b>	<b>140,000</b>	<b>140,000</b>	<b>600,000</b>	
<b>Total cost across five years</b>	<b>1,885,000</b>						
<b>Estimated cost per year (spread over 5 years)</b>	<b>377,000</b>						

UNESCO Institute for Statistics –Options for Global Learning Metric

<b>Projected budget for ULS Stream 2 activities cost for Alternative 3 (A new assessment)</b>							
<b>Activity: Other Professional Services</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>Total</b>	<b>Notes</b>
Consultation with regional assessment bodies	35,000	35,000	35,000	35,000	35,000	175,000	Online platform and/or in person consultations.
Consultations	20,000	20,000	20,000	20,000	20,000	100,000	Specific topic consultation with advisory board members
Research cost	0	100,000	0	0	0	100,000	Appointed research or commissioned papers
Develop field data collection procedure cost	300,000	0	0	0	0	300,000	International cost, assume not totally working from beginning given the experience in this area. Invest in developing new procedures, depending on assessment design.
Translate standards/guidelines in major languages		100,000	0	0	0	100,000	International cost - translation cost on guidelines documents
Country field test data collection cost	0	0	2,000,000	4,000,000	0	6,000,000	Country implementation cost to test framework of methodologies 1-2 countries per (5 or less) regions. Maximum 10 countries. First year is the FT and second year is the MS.
Translation cost for documents	40,000	40,000	40,000	40,000	40,000	200,000	Translating documents to 4 other languages, if necessary (French, Spanish, Arabic, Russian)
<b>Total</b>	<b>395,000</b>	<b>295,000</b>	<b>2,095,000</b>	<b>4,095,000</b>	<b>95,000</b>	<b>6,975,000</b>	
<b>Activity: Travel, Conferences &amp; Meetings</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>Total</b>	
Staff Travel	20,000	20,000	20,000	20,000	20,000	100,000	Staff travel for additional consultation; travel to SC or TAG meetings if meetings not in Montreal
Conferences	10,000	10,000	10,000	10,000	10,000	50,000	Presentation at conference to rise visibility and awareness of the project
Steering Committee (SC) meetings	20,000	20,000	20,000	20,000	20,000	100,000	Organizing of SC meetings once a year
Technical Advisory Group (TAG) meetings	40,000	40,000	40,000	40,000	40,000	200,000	Organizing of TAG meetings once a year
Implementation Training Workshops	0	0	100,000	50,000	0	150,000	3 workshops (USD 50,000 each), include regional bodies and participating countries. # workshops depend on the feasibility studies outcomes.
<b>Total</b>	<b>90,000</b>	<b>90,000</b>	<b>190,000</b>	<b>140,000</b>	<b>90,000</b>	<b>600,000</b>	
<b>Total cost across five years</b>	<b>7,575,000</b>						
<b>Estimated cost per year (spread over 5 years)</b>	<b>1,515,000</b>						

UNESCO Institute for Statistics –Options for Global Learning Metric

Summary total cost, political aspect and additional notes

<b>Total cost</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>total</b>	<b>3-year cost</b>
<b>Staff cost</b>	282,052	282,052	282,052	282,052	282,052	<b>1,410,260</b>	846,156
<b>Alternative 1: Backward Linking</b>	285,000	4,585,000	235,000	0	0	<b>5,105,000</b>	
<b>Alternative 1.1: Forward Linking using CNAs</b>	285,000	2,585,000	235,000	0	0	<b>3,105,000</b>	
<b>Alternative 1.2: Forward Linking using NLAs</b>	735,000	2,585,000	235,000	0	0	<b>3,555,000</b>	
<b>Alternative 2: Adapt an existing test</b>	635,000	545,000	235,000	235,000	235,000	<b>1,885,000</b>	
<b>Alternative 3: Develop a new test</b>	485,000	385,000	2,285,000	4,235,000	185,000	<b>7,575,000</b>	
<b>Alternative 1: Backward Linking</b>		<b>5,951,156</b>	Note for Alternatives				
<b>Alternative 1.1: Forward Linking using Cross-National Assessments</b>		<b>3,951,156</b>	1, 1.1, 1.2 Staff cost is				
<b>Alternative 1.2: Forward Linking using National Learning Assessmen</b>		<b>4,401,156</b>	for 3 years				
<b>Alternative 2: Adapt an existing test</b>		<b>3,295,260</b>	These two alternatives				
<b>Alternative 3: Develop a new test</b>		<b>8,985,260</b>	will be for 5 years.				

**Appendix H – Costing Regional Modules**

Cost for a regional module using PISA for development as the reference assessment

Table 1. Test development costs

Task area description	Total estimated costs (USD)
Adaptation of cognitive items from item pool for the region/language	25,000
Equating study	25,000
Develop background information based on regional needs	15,000
Training materials development (best practices manual, interviewers manual, field test administration, case management guidelines), adapt from reference assessment	25,000
Field test in a few selected countries on the instruments and procedure in the region (3X40,000)	120,000
Database preparation and analysis	25,000
Management costs	100,000
Total cost	335,000



Table 2. Cost for each additional language of cognitive instrument

Description additional language costs	
Translation/verification of cognitive instrument to another language if different from source language (cost per language)	10,000
Layout and assembly of the assessment in new language(s)	10,000
Scaling and analysis	10,000
Total cost	30,000

DRAFT

### ***Appendix I. Learning Metric Partnership (LMP) and Global Alliance for Learning (GAL)***

The Global Alliance for Learning is the platform that brings together all key actors and components (methodological development, capacity-building and funding) needed to help countries achieve SDG4 and leverage the use assessments to improve the quality of learning outcomes for all. The Partnership will build on existing initiatives, while fostering collaboration across assessment agencies, implementers, national governments, civil society groups and the international education community.

The Global Alliance for Learning (GAL) is uniquely designed to:

1. Become the platform for developing the common measurement framework for SDG4. The GPL would oversee the definition of indicators and development, over time, of the instruments to measure progress toward SDG4. Specifically, it would:
  - Establish and maintain a Catalogue of Learning Assessments
  - Develop and implement a Global Snapshot of Learning Outcomes
  - Develop a common framework for Data Assessment Quality that informs and guides efforts by all countries
2. Coordinate a global communication campaign to build the sustained support for the efforts to collect and use high quality information on learning outcomes in all countries around the world. Specifically, it would:
  - Organize global consultations to promote buy-in for the agenda
  - Make information available to all stakeholders in a creative and user-friendly fashion, etc.
  - Share best practices across the world
3. Support developing countries in their efforts to build stronger learning assessment systems that inform policy and programs at the country level. Specifically, it would:
  - Facilitate the preparation of capacity building plans at the country level
  - Provide a cost-effective platform to link developing countries with organizations providing technical assistance and support
4. Mobilize the resources necessary to support the development of stronger learning assessment systems and the collection of the data needed to monitor SDG 4. Specifically, it would:
  - Lead the process of establishing a funding mechanism for capacity building of learning assessment systems
  - Help to ensure that investments needed to collect the data to monitor SDG4 is secured

On the other hands, Learning Metric Partnership (LMP) is a program aims to develop a set of nationally and internationally comparable learning metrics in mathematics and reading, and then to facilitate and support their use for monitoring purposes, in partnership with interested countries. The key features of the initiative are fourfold:

- It accommodates results from a range of different assessments of learning outcomes.
- It yields high quality data that are nationally relevant and internationally comparable.

- It emphasises peer-to-peer capacity support and learning opportunities.
- It has a strong focus on improving data use and policy development.

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

In summary, the main difference between GAL and LMP is that GAL is the platform to bring stakeholders together while LMP is the program that develops the metric. LMP will be subsumed under GAL as it is one of the programme that produces indicators for SDG4. While GAL covers the whole spectrum of activities for SDG, LMP is just one of the programme that produces relevant indicators for global monitoring.

***Appendix J: Notes on paired comparison***

Assuming that paired comparison means comparison between two independent assessments. Before any comparison could be made the two assessments will need to align onto a common baseline so comparison could be meaningful. To be more specific it is assumed that the paired comparison is to compare the mean scores adjusted by variation between two independent assessments.

First, the score metric is usually not the same across national assessments or two independent cross-national assessments, so alignment will be needed for the two comparing metrics. In other words, one score metric will need to be put onto another so that they are on the same platform. For example, if Assessment A has a score range of 0-100 and the mean score is 50 and Assessment B has a score range of 200-500 and the mean score is 350. It is therefore not reasonable to just take the difference of the two scores; adjustment of the scores is required before comparison.

Second, due to the range of scores the variabilities of each of the two assessments will be difference. In addition, the reliability of each assessment will be difference so again the variances will need to be adjusted.

Paired comparison is usually done within variables of cross-national assessment, e.g. the comparison of the performance between male and female; those who live in urban versus rural, and those with high SES versus low SES. The comparison is meaningful since these are subgroups within the assessment.

Comparison across countries within the same cross-country assessment, like PIRLS, is also feasible since countries who participated in PIRLS have been scaled onto the same common metric.

When comparing assessment results of countries who participated in two different cross-national assessments, like country participated in TIMSS and country participated in PISA (Mathematics) statistical adjustment will need to be performed so that the comparison will be meaningful.