Draft Summary Report

Technical Expert Meeting on SDG Indicator 4.6.1

Paris, 7–8 November 2017

Background and objectives of the meeting

This expert meeting was organized by the UNESCO Institute for Statistics (UIS), in collaboration with the UNESCO Institute for Lifelong Learning (UIL), within the framework of the Global Alliance to Monitor Learning (GAML). Its aim was to support and inform the ongoing work of the Task Force on Sustainable Development Goal (SDG) indicator 4.6.1 on youth and adult literacy and numeracy.

The meeting brought together academia, international and national assessment specialists, different UNESCO entities and international development agencies. These experts convened in order to enrich discussion from their different perspectives on developing a methodological framework and strategy for measuring Target 4.6, as outlined in the box below.

SDG Target 4.6: By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy.

SDG Indicator 4.6.1: The percentage of the population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills.

The objectives of the meeting were to:

- Reach an agreement on an 'expanded' conceptual framework for indicator 4.6.1, including domains for the global assessment framework for indicator 4.6.1
- Explore the existing framework and/or propose the inclusion of alternate ways of measurement of the identified relevant competencies for indicator 4.6.1
- Suggest a pragmatic action to define a minimum or fixed level of proficiency in the domains of literacy and numeracy on indicator 4.6.1.

Day 1: Presentations and discussions

In the opening session, Margarete Sachs-Israel, Chief Programme Coordinator, UIL presented the objectives of the meeting. It was also pointed out that the aim of the expert meeting was to inform the work of the Task Force on 4.6.1 and to provide pragmatic recommendations for the measurement strategy for 4.6.1.

Session 2: Background and working framework – Challenges and opportunities to report on indicator 4.6.1

Silvia Montoya, Director of UIS, in a presentation entitled 'Challenges and opportunities to report on the indicator 4.6.1', stressed the importance of moving quickly given the tight deadlines that must be met in order to monitor this target effectively by the end of 2018. As the custodian of the SDG 4 indicators, UIS is required to update the Technical Cooperation Group (TCG), which will, in turn, report to the Education 2030 Steering Committee on the status of each indicator, including 4.6.1, in December 2017. More importantly, there was a need to develop proposals for the Tier III indicators by December 2018, including estimates of the cost of delivery.

Ms. Montoya briefed participants about the current status of data sets that could be useful in reporting against indicator 4.6.1. Those data sets are generated through both indirect and direct assessments conducted at global and national levels. There are a few cross-national assessment surveys, such as the Organisation for Economic Co-operation and Development's PIACC (Programme for the International Assessment of Adult Competencies), the World Bank's STEP Skills Measurement Programme, and UNESCO's Literacy Assessment and Monitoring Programme (LAMP), which are largely based on the same conceptual frameworks. While a majority of countries in the world still use a dichotomous definition of literacy (i.e. literate or not), a number of countries have undertaken their own national assessments to monitor the literacy skills of their adult population. Two of these countries – Kenya and Bangladesh – were represented during the meeting. There was, however, an issue concerning differences in constructs and frameworks adopted by different national assessments, which might pose challenges in cross-national comparisons.

Ms. Montoya highlighted three key questions: 1) What and who to assess; 2) how to assess; and 3) how to report. She noted that answering these questions demanded discussion of definition and content, methodological and operational frameworks, and reporting levels and periodicity.

Participants were also provided with a summary of the main definitions currently in use by different assessments, sourced from UIS metadata. For example, the UNESCO definition[1] was employed in two UNESCO-led literacy assessments, LAMP and RAMAA. Other initiatives, such as the European Literacy Policy Network: European Declaration of the Right to Literacy, PIACC and the STEP Skills Measurement Programme defined functional literacy and numeracy in a similar but different manner.

[1] 'Literacy is the ability to identify, understand, interpret, communicate and compute, using printed and written materials associated with varying contexts. It involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society'.

Ms. Montoya also presented criteria to measure functional literacy and numeracy, including: a) a definition that invokes a continuum, b) an assessment which covers a full range of skills, c) statistical methods which confirm psychometric stability, and d) statistical methods to support cross-national comparisons.

Using this set of criteria, Ms Montoya provided an analysis of the existing direct and indirect cross-national and national assessments. National indirect assessments which a large number of countries currently conduct, conceptualize literacy as a dichotomy, either you are literate or illiterate, rather than as a continuum, having a range of skills. This information collected through different surveys conducted at country were used to report on UIS literacy rate. Neither of these assessments meet the criteria for measuring 4.6.1. On the other hand, cross-national direct assessments, such as PIACC, IALS, ALL, STEP and LAMP, meet all of the four criteria, while RAMAA, implemented in 12 countries in Africa, meets only one out of the four criteria. With regard to direct national assessments of literacy, only Canada met all the criteria. The USA, New Zealand, India and Kenya met at least three out of the four criteria.

Ms. Montoya highlighted the challenges in using different assessments to report on the same indicator. For example, while both PIACC and LAMP evolved from IALS and ALL, they differ in terms of content though with some common items, mode of delivery and reporting methodologies. However, they evolve from the same IALS/ALL framework and there are enough commonalities to allow harmonization and statistical comparison through model-based estimation.

Regarding reporting on the SDG 4, Ms. Montoya stressed the importance of pragmatism and balancing the need for interim reporting from 2017 with available data and tools; and developing a methodology for sustainable reporting in the long run.

For the long-term monitoring of Target 4.6, there is a need to adopt, adapt or develop a new tool to measure progress, which embodies equity and is accessible by all as a global public good. Also needed are key principles to build on existing work, the use of national data, where possible, and the use of non-official data to fill the gaps while balancing quality with fitness for purpose. To set such criteria, the following aspects must be considered: whether a measurement framework sufficiently covers required domains; the properties of the tool; and the properties of data for reporting.

It is also critical to agree on the essentials of comparability, linking and/or anchoring methodologies to maximize coverage to measure progress against SDG 4. It is also critically important to propose the interim reporting strategy (IRS), while linking methodologies and tools are being finalized, that will provide progress into long-term sustainable reporting strategy.

UIS further proposed the following steps to move forward:

1. To determine an initial list of criteria for using existing data with specific attention to instrument that use to produce the data and quality of data.

2. To identify further methodological work to be conducted, including examine methodology that could link existing datasets, as well as developing new methodology, by convening a group of experts to set priorities and compare different methodologies.
3. To decide on a pragmatic and workable framework for interim and long-term reporting.
4. To consider the alignment of skills with basic education.

Ms. Montoya closed the presentation by showing a possible way forward for interim reporting, using national data with appropriate footnotes describing the quality and comparability of the data.

Session 3: Frameworks of the existing cross-national and national literacy assessments
The existing cross-national and national direct literacy assessments of PIACC, STEP and LEO (Level One) were presented.

OECD's PIAAC Literacy and Numeracy Frameworks

William Thorn, Senior Analyst, OECD, started with a reference to Ms Montoya's presentation and the challenges that lie ahead. He provided a brief overview of the PIAAC assessment framework and its likely development in its second cycle and beyond. The PIACC assessment framework defines the features of the construct to be measured, and guides the development of test items as well as the interpretation of results. Following the first cycle, the second cycle is about to start (2018–2023).

PIAAC's defined literacy as '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'. It has the following characteristics:

- 'Use' oriented concept of competency (i.e. adults read, deal with mathematical information to do things in the world).
- A continuum of proficiency levels.
- Covers both cognitive and behavioural dimensions.
- Is a socially purposeful activity.

Three main dimensions of the construct are:

1) Content (the artifacts, tools, knowledge, representations and cognitive challenges that constitute the corpus to which adults must respond or use). This includes print-based or digital types of texts in different formats, such as continuous or prose texts, non-continuous texts, mixed texts and multiple texts.

2) Cognitive strategies (the process that adults must bring into play to respond to or use given content in an appropriate manner). Cognitive strategies involve functions such as accessing and identifying text, integration and interpretation, and evaluation and reflection.

3) Context (the different situations in which adults have to read, display numerate behaviour and solve problems. The texts in tasks represent personal contexts, society and community.. PIACC assessed the respondents' response to written text (reading), not the production of written text (writing). The numeracy framework adopts a similar approach.

In the second cycle of PIAAC, while minimal change has been made to the literacy component, a significant change has been introduced to numeracy. It now includes the ability to engage with digital representation and tools. Numeracy components might be developed to identify people with low numeracy levels. It was noted that considerable commonality exists across frameworks in different studies, such as IALS, PISA, PIRLS, TIMSS and PASEC.

An assessment framework should aim to define a construct that is relatively context independent (i.e. relevant across different population groups and over time. There may be other assessment frameworks of literacies or numeracies,but it is difficult to conceive constructs that do not have considerable commonalities with those of PIAAC.

World Bank Group's STEP Household Survey: Operationalizing the PIAAC framework in middle-income countries

Koji Miyamoto, Senior Economist, World Bank Group, explained that, built on the PIACC framework, STEP was conducted in 15 countries and mainly among urban populations in middle-income countries, including Kenya. Currently, preparation is underway for Afghanistan to participate in STEP. The sample size of the STEP household survey is 2,000–3,500, targeting adults between 15 and 64 years old. It is a paper-pencil assessment which takes about 2–2.5 hours to complete. Countries can opt for a partial or full module. In addition to the test assessment, household and individual information was also collected. There was good reason for thinking PIACC could work in middle-income countries, however it was financially and administratively demanding for many countries.

Existing national framework to improve measuring of lower literate population – The case of Germany's Level One Survey: Theory and Item Construction

Anke Grötluschen, Professor, the University of Hamburg, Germany, presented the framework of the Level One Survey (LEO) from Germany. Based on the British Skills for Life Survey, the LEO Survey attempts to provide a fair estimation of functional literacy with a particular focus on the low end of the ladder. LEO uses the UNESCO definition of literacy to assess reading and writing and its proficiency levels are theoretically linked to PIAAC levels.

LEO has five 'alpha levels' ranging from Alpha Level 1 (reading at letter level) to Alpha Level 5 (reading at text level), but Alpha Level 3 is still below PIACC Level 1. Each alpha level is described by difficulty-determining factors and can-do descriptions. Anke Grötluschen also informed the meeting of Germany's plan to link PIACC and LEO statistically.

There are considerable methodological differences between PIACC and LEO. Ms Grötluschen described the differences and why, ultimately, LEO is based on the British approach.

LEO uses two-step approach, comprising constructing reading (decoding/recording) and lexical reading (lexical entry of the word into the mind). It considers LIX (index of readability), including length of words and sentences, complexity of sentences, cluster of consonants, word frequency, and typography.

Considering the transferability of the assessment to other languages and regions, she said that reading was much more transferable than writing.

Discussion

Several issues were identified as critical in the existing assessment frameworks and domains.

- Impact and use of surveys: A key consideration is the usefulness of assessments to countries. In this context it was asked how these surveys contributed to policy formulation, what their value was for countries, and what we could learn from them going forward.
- Limitations of STEP: Concerns were raised about the limitation of STEP in urban areas and the importance of national needs when collecting data. Koji reported that in some countries the survey was extended to rural areas, which is very important for covering 4.6.1.
- Domains: All the definitions of literacy refer to writing as an integral part of adult literacy skills, especially in our increasingly digitalized world. However, it was generally agreed that writing is more difficult to measure and assess and requires the evaluation of psychometricians in scored responses and complex modelling in scaling. In addition, trained scorers and complex operation in scoring to evaluate responses. Writing could, however, be part of national surveys as indicated in several examples (e.g. LEO in Germany; Korean Literacy Assessment; IVQ in France; NAPLAN in Australia). There was a broadly shared sense, however, that global writing assessment with cross-country comparability would not be recommended, due, for instance, to linguistic comparability, including different levels of transparency in orthography. To capture lower-end skills, national sub-tests for writing skills could be conducted.
- Participants agreed that assessments should test more than the ability to deal with familiar contexts which can bias the results of tests. Increasingly, there is a need to assess the ability to deal with the unfamiliar. The issue, however, is how to make items that are unfamiliar to all.
- Time: The time involved in each survey needs to be considered. STEP, for example, takes 2.5 hours including the background questionnaire. This is an important consideration not only for a stand-alone test, but also if one would like to implement a short literacy module to existing household surveys.

- Costs: The cost of surveys aiso a key consideration and can often prevent countries from participating in cross-national surveys or even undertaking their own. STEP, for example, costs 500,000 USD per country.
- Participants also learned about OECD's plan to develop a short literacy survey (to be administered on a tablet), which could be of used for SDG 4.6.1 monitoring. The first step is to develop a test, scale the test and translate versions in different languages.

Contextualization of cross-national assessments

Bryan Maddox, Senior Lecturer, University of East Anglia, U.K., presented some of the findings of his ethnographic research on UNESCO's LAMP experience at country level, focusing especially on adult literacy assessment events in rural Mongolia. He provided several illustrative examples of the issues of contextualization or the lack of it when global literacy assessment was conducted at national levels. The issues of 'too much contextualization' and 'not enough contextualization' was explored, based on observations made regarding the relationship between textual and contextual knowledge of a respondent when Mongolian nomadic herders participated in the LAMP assessment. The problem was observed when the assessment content was too distant from the lived reality of the respondent. When test items are 'too much contextualized', however, it was also problematic as the respondent privileged contextual knowledge linked to his real life over textual knowledge contained in the assessment, which should be the reverse. Mr Maddox also pointed out the potential influence of socio-cultural factors on the assessment process and outcomes, as was the case of the Mongolian nomadic respondent whose family was eager to help the respondent as part of their culture of collaboration. The challenge is to find a point of balance in contextualization. Maddox stressed the importance of testing the validity of assessment in local contexts as formal validation standard processes might also mask the differences in performance. Therefore, different validation standards should be identified. He also highlighted key issues to be further explored, such as resources and support, capacity, adaptation, gender bias and the cultural fit of the assessment.

Participant agreed that:

- In using large-scale assessment across countries, more feasibility studies are required to understand the cultural adaptations. As there were serious policy and financial implications of assessments, sufficient time is needed to produce a good contextual fit and to eliminate errors in the assessment. However, the ambitious plans and timelines for reporting on indicator 4.6.1 will make such studies difficult.
- The question of contexualization was considered very important and interesting. It might be helpful to consider what items can be decontexualized. Looking at how national surveys deal with the contextualization issue could also be useful.

- Another idea was to create a large pool of items through adaptive testing, although its use in low- and middle-income countries and the complexity of creating such a large pool of items could be a challenge.

Session 5: National literacy assessments: Cases of the Bangladesh Literacy Survey and the Kenyan Adult Literacy Survey

During this session, Kenya and Bangladesh described how they had implemented literacy assessments.

Kenya National Adult Literacy Survey (KNALS)

Janet Chepkemoi Rotich, Deputy Director, Directorate of Adult and Continuing Education, Kenya, explained that the **Kenya National Adult Literacy Survey** (KNALS) was carried out to fill in the gaps for informed policy-making. Following the first survey in 1988, the second survey, KNALS-I, was conducted in 2007. Since 2016, Kenya has been developing KNALS-II, based on UNESCO's operational definition of literacy. Although it was initially planned to adapt UIS-LAMP for the second KNALS-1, it was realized that LAMP's objectives were different from those of KNALS-I, which aims to establish baseline information on adult literacy levels as well as to obtain comprehensive data from programme providers.

Eventually, it adapted both Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and LAMP models. In KNALS-I, a nationally representative sample of 18,000 households was included. Literacy was measured both directly and indirectly (through self-declaration).

Tests were translated from English into 19 main languages and the survey was conducted in 20 languages. The survey questionnaire included a household questionnaire, an individual questionnaire, an institutional questionnaire, and a direct literacy assessment.

Performance of youth and adults was measured based on continuum of skills. Literacy components focused on narrative prose, expository prose (continuous text) and documents (non-continuous text), while the numeracy component included number, measurement and space. The range of the scale was from Level 0 (cannot read or write) to Level 5 (can read forward and backward through a text in order to confirm understanding). Level 3 applied to a literate person who had acquired the essential basic knowledge and skills of reading and numeracy and was able to sustain literacy skills. A person who attained Level 4 was considered as having the desirable cognitive skills.

The findings of KNALS-I revealed the geographic dimensions of literacy skills distribution across the country with Nairobi province having above-the-national-average literacy competencies, with north eastern province scoring the lowest scores. Nationally, 29.6 per cent attained the desired literacy competency levels and 61.5 per cent attained the minimum mastery levels. In the second phase of KNALS, there would be eight competency levels, as opposed to five in the KNALS I. Unlike the previous assessment survey, KNALS II would have an explicit assessment

framework. The benchmarks would be aligned with the Kenyan National Qualification Framework and Curriculum for Formal and Non-Formal Adult education.

Literacy Assessment Surveys in Bangladesh

A.K.M. Ashraful Haque, Project Director, Bureau of Statistics, Bangladesh, presented on the **Literacy Assessment Survey 2011** and the plans for the **Literacy Assessment Survey in Bangladesh in 2018**. Mr. Haque gave an overview of Bangladesh as a country, followed by a presentation on literacy assessment in Bangladesh. Bangladesh has already produced data for indicator 4.6.1 within the framework of SDG 4. In Bangladesh, the definition of literacy used in censuses between 1961 and 2011 by the Bangladesh Bureau of Statistics was based on the decision that a person of 7 years and older who was able to read and write a letter should be considered literate.

Mr. Haque explained that Bangladesh will adopt UNESCO's definition of literacy. In Literacy Assessment Survey in 2011, the test instrument measured four domains: reading, writing, numeracy, and communication and comprehension. Four competency levels were identified: non-literate (lack ability to recognize and write alphabets, words, and numbers, score between the range of 0 and 24.99), semi-literate (ability to recognize and write some simple words, to count objects and numbers at a very basic level, scoring in the range of 25 and 49.99), literate at initial level (ability to read and write sentences, possessing skills of four basic rules of arithmetic and limited use of literacy skills, scoring between 50 and 74.99) and literate at advanced level (ability to read and write fluently, competency in four arithmetic rules, ability to use the skills in everyday life, scoring between 75 and 100).

Discussion

- Ms. Rotich explained that Kenya used a large sample size in order to cover 47 counties. She also confirmed that Kenya used Blooms' taxonomy to define the different levels of skills and that they planned to expand the levels.
- While both countries used UNESCO's definition of literacy, participants noted that they measured different domains, have different proficiency levels and cut-off points which would make comparability challenging.
- Participants also discussed the challenges associated with languages and translation. In the case of Kenya, rigorous methods were employed to ensure the meaning remains unchanged by doing both forward and backward translations and the employment of three translators per language.
- Other questions concerned methods to ensure item difficulty levels will be maintain through translation, and the need to have assessment framework for both literacy and numeracy. In Kenya, the assessment framework of KNALS-II has been developed and aligned with the Kenyan National Qualification Framework and Curriculum for Formal and Non-Formal Adult

education, which facilitates benchmarking progress and ensure equivalencies of competence levels across formal, non-formal and informal pathways.

Session 6: Summary of Day 1

The main conclusions and agreements reached at the end of Day 1 were as follows:

- It was agreed to adopt the UNESCO definition of literacy (2004) as the conceptual framework since it does not eliminate measurement issues, does not clash with other definitions and is currently in use by many countries.
- Participants also discussed the importance of linking the definition with the levels of proficiency and differentiating between a definition, an assessment framework and an assessment tool, as these are different things.
- Regarding domains, it was agreed that the global assessment framework should focus on reading and numeracy. Writing, which was considered important, could be assessed at the national level if countries so desired.
- Substantial variations in proficiency levels across national and international assessments posed the challenge of global comparison but the existing dataset included a wide variety of domains and sub-domains to assess literacy and numeracy of the adult population for different purposes.
- The challenges concern how to define levels, as well as different domains and sub-domains, when skills are assessed for different purposes. It was also recognized that diverse proficiency levels could pose a challenge for global comparison.

Day 2 sessions and discussions

Session 7: A synthesized numeracy framework

Assessing adult numeracy: In search of a framework for sustainable monitoring

Iddo Gal, Senior Lecturer, University of Haifa, Israel, presented a paper produced for the Global Education Monitoring Report (GEMR) regarding an assessment framework for the sustainable monitoring of adult numeracy. Not much conceptualization has taken place regarding numeracy. The main question related to literacy and numeracy and whether numeracy should be subsumed in literacy or be treated separately. Considering UNESCO's definition of literacy, in which numeracy was subsumed under literacy, he proposed ways to reduce the impact of literacy skills on numeracy assessments.

Mr. Gal contrasted the definition of numeracy skills from UNESCO, IALS, ALL, PIAAC, LAMP, Kenya and Bangladesh. He proposed that the PIACC definition and extension of this definition to the lower end should be informed by the experiences and results of LAMP, LEO1, and other national assessments. Regarding the reporting framework for numeracy, he recommended increasing sensitivity at lower end of the scale. Proficiency levels could be based on PIAAC's conceptual framework, which is on a continuum of five levels of proficiency with the extension of numeracy components as the Level 1 tasks require the respondent to show an understanding

of basic numerical ideas by completing simple tasks in concrete, familiar contexts, where the mathematical content was explicit with little text. Level 1 could be items to cover rudimentary numeracy skills, which would be tested orally with minimal dependence on reading.

Regarding a numeracy assessment framework, Iddo Gal recommended that a test with 16–20 items be developed by international teams and countries, for which three options were proposed: Option 1: single test with 16–20 items, covering all five levels, total score based on number correct (no IRT), used by all countries, with cultural adaptations; Option 2: Single test with 16–20 items, covering 3–4 levels for countries to choose from); and Option 3: more sophisticated test (IRT-based), rotating booklet design with all the implications.

Discussion

The following questions and comments were made during the discussion.

- Regarding measurements of numeracy, which are emerging as assessments distinct from literacy ones, there are the potential needs of such data. In Germany, people are interested in numeracy data, for instance, for refugees, people with special needs, and STEM professionals. In Israel, although policy-makers have never seen the numeracy data, their future decision-making would require this type of data. For PIAAC, the focus, for instance, is on improving literacy not at individual level but at population level. The numeracy data is required for SDG.4.6.1 monitoring as target and indicator include numeracy.
- For greater robustness of numeracy assessments, rigorous empirical testing is required. To design and implement assessments and roll them out at country level, further discussion and reflection are required where there are issues related to capacities, costs and time constraints. For instance, establishing a minimum proficiency, which differs across countries, would take a long time.
- Level 1 study could be recommended to countries, especially low income and lower middle-income countries, to then link it with international and regional tests.
- Considering the lower-end of the skills continuum, PIAAC has items for the lower end for literacy that include numeracy. But the question is how much information is really needed for effective policy-making. It is critical to decide on the need to get more information using different methods, orally or digitally, for the lower end, whether through a diagnostic or component skill survey to get the profile of low literate population. IVQ's model for lower-end literacy skills uses oral questions in numeracy, to which answers can be provided in writing or orally.
- One broad question to ask is what policy-makers want out of the SDG 4 monitoring processes and whether the lower-end proficiency levels are really necessary. Participants proposed a UNESCO-led survey to assess the needs of policy-makers in terms of the use of SDG 4 from their perspectives.
- UNESCO stressed the importance of considering the equity aspects of SDG 4, which requires information on populations typically at the lower end of proficiency levels.

Session 8: Coherent perspective on conceptual framework

Definitions of literacy and numeracy, and a strategy for measurement

Scott Murray, President, DataAngel Inc., presented a strategy for the measurement of indicator 4.6.1. Mr Murray presented the strengths and weaknesses of PIAAC. A positive aspect of PIAAC is that it provides valid, reliable, comparable and interpretable results. However, implementation errors and technical and operational demands can result in catastrophic results making the survey even more costly than its price.

Mr. Murray described the case of Canada and proposed that any country which has the funds and implementation rigour can and should go with PIAAC. He also stressed the importance of translating the results into meaningful policy changes.

PIAAC results should be systematically linked to efforts to improve teaching and learning as in Canada, and international test results must be translated into instructional improvements if progress toward the SDGs is to be achieved. Mr Murray stressed that functionality can only be decided based on the purpose or intent of the campaign. Functionality could only be defined relative to whose interests were being served. This required further conceptual reflection, including topics such as how literacy can help individuals to cope with change or to introduce change, the skills that individuals need to impact economic, fiscal and tax investments, and how to equip individuals with skills to maximize the macroeconomic performance of the country in a more equal manner. Murray highlighted that the proficiency levels should not be imposed on countries and that it should be left to countries to decide on which levels meet their own objectives.

Based on these observations and his experience in the field of national and cross-national literacy assessments, Mr Murray recommended the following:

1) OECD countries to continue using PIAAC.
2) Countries close to the OECD average and having the means of implementation to implement PIAAC.
3) Other countries to:
    a. adopt fully adaptive, computer-based testing;
    b. improve the precision of proficiency estimates by adopting fully adaptive, computer-based testing and validation;
    c. reduce financial and operational demands by applying purposive sampling and synthetic estimation;
    d. decide on their own proficiency level.

In order to drive the policy agenda and generate interest, Mr Murray proposed that synthetic or model-based estimates be generated for each country using the best available data. Finally, he

described some exciting technological advances which could be implemented easily in technology- or internet-poor areas at relatively low cost.

Outcomes and agreements reached

- For policy-makers to make sensible decisions, the scientific data on skills levels is useful. Murray mentioned that, in Canada, the massive skills loss detected by the statistics induced a policy shift. When technology is increasingly taking over the jobs that can be performed by low-skilled adults, policy-makers need to think about how to increase the number of people at skills levels 3–4 and above. Although this might not happen within 10–15 years, it is a possible change to anticipate.
- Regarding monitoring for SDG 4.6.1, the importance of realism was reiterated. To produce data from a mix of existing assessments within one year is impossible. Caution was also sounded regarding the lack of reliable census data, the need for huge infrastructure and resources, and challenges related to the linguistic and cultural dimensions of measurements. At the same time, the need to anticipate technological advancement in managing handling measurements was pointed out.
- Participants agreed that the financial and operational demands could be reduced. In any event, an assessment is cheaper than a jet fighter.
- It was also questioned to what extent we want to associate global monitoring efforts with a neo-liberal rationale (linked with testing companies).

Session 9: Pragmatic approach to measuring literacy and numeracy to improve the coverage of indicator 4.6.1

<u>Challenges and opportunities to report on the indicator 4.6.1: Closing remarks of UIS</u>

Ms. Montoya reiterated the objectives of the meeting that needed to agree on a definition of literacy, a measurement framework, measurement tools, and interim and long-term reporting. She elaborated on the following elements when considering the measurement and monitoring on indicator 4.6.1:

1. **Definitions**: There is agreement that the UNESCO definition of literacy will be used for the measurement of indicator 4.6.1.

2. **Measurement framework**: PIAAC has been adopted. More work is needed to capture the lower end of the continuum by this framework (or others) and to map other surveys against it.

3. **Measurement tool**: Much of the discussion focused on assessment tools. A range of different tools, including those that cover the lower end of the literacy and numeracy skills continuum, should be considered. How proficiency scales or performance levels of

various surveys are/can be aligned? What are the desirable properties of the tool? We have different tools, but the framework should allow these tools to define a wide range of skill proficiencies that correspond to the proficiency scale of the framework. Different tools also mean different cost and operational demands.

4. **Interim and long-term reporting**: In terms of reporting, for the short run we need to be pragmatic, but also rigorous, as needed, and with a long-term view. We also need to be as open as possible to measure progress. Ideal criteria for reporting could include the following considerations:
    - A definition that is agreed and accepted.
    - A measurement framework that covers necessary domains and sub-domains.
    - Meeting the criteria for the acceptable properties of the tool and data.
    - Is there any way for comparable reporting?

5. The period of reporting should be decided and recommended to UIS. The GAML Task Force on indicator 4.1.1 agreed, for instance, the reporting period of 2010–015 and 2016–2020.

6. The issue is the slim data coverage on indicator 4.6.1. There is a need to come up with criteria for data and measures. This includes the qualities of instruments and databases that are necessary/critical and desirable. Data-sharing parameters should be determined. Further methodological work should be carried out, including convening a group of experts to specify the above-mentioned and compare various methodologies, as well as to identify the need for further data collection.

7. The following practical steps could be considered:
    - **Definition of content standards** in relation to domains and sub-domains: using existing framework or not, mapping cross-national literacy and skills survey frameworks to identify gaps and define the expansion, and mapping skills survey frameworks, curriculum and national qualifications frameworks, and proposing a draft assessment framework of reference. UNESCO IBE's curriculum mapping could be used as a reference list.
    - **Performance levels:** National, regional and global proficiency levels.
    - **Proficiency scales**: Proficiency scale map could be developed to explore how proficiency scales or performance levels of different surveys are/can be aligned.
    - **Conceptual moderation of proficiency levels**: What should be test scores of youth and adults would be then classified into the desired performance level for the indicator 4.6.1 reporting.

Ms. Montoya further suggested the following possible and practical steps to consider advancing further the development of the measurement for indicator 4.6.1:

- Define common domains and sub-domains, continuum of skills.
- Define number of skills levels and categorize the continuum of skills.
- Write policy descriptors (generic term) for the defined levels.
- Develop full descriptions for the performance levels of the UIS reporting continuum.
- Choose a global or regional reference level of functional literacy and numeracy as the fixed proficiency level for reporting.

Discussion and agreements reached

- Regarding measurement tools, it was proposed to explore the potential of OECD's short literacy survey (SLS) currently under development for SDG 4.6.1 monitoring. It is a relatively short instrument with 1 to 3 levels on a tablet with the capacity for automatic scoring. It covers literacy only, not numeracy. It can be a standalone/dedicated study or introduced as a module in an existing international household base survey such as the Multiple Indicators Cluster Survey (MICS) or the Demographic and Health Survey (DHS). The progress has been made as the test is developed in English but it needs to be validated through field testing. The next step is to check its links with the PIAAC scale and pilot in additional languages. The SLS tools is available and downloadable. Since it has not been fully validated people will have to use it at their own risk.
- In defining performance levels, two interrelated aspects should be considered. On the one hand, performance levels are used to categorize the ability continuum on a scale in assessment. On the other, performance levels are related to 'standards' and 'norms' external to the tests. The scales used by PIAAC and PISA are data-driven, defined by the groups of countries that participated in the assessment. It is important to rethink the process.
- In addition, it was recommended not to start with descriptions of performance levels. It is worth checking into some existing surveys (LEO) how they describe levels and it is usually at the later stage of the assessment process with information on the performance of items. The development of IVQ started with research on existing work and skills levels and adjusted descriptions of skills levels after eliminating items not conforming with the scale, using IRT.
- On the Expert Group, it was questioned how this Expert Group would proceed, with what kind of timeframe and which terms of reference?

10. Summary of agreements reached, recommendations and closing remarks

The meeting was concluded by Margarete Sachs-Israel reiterating the agreements reached during this meeting, as follows:

- Adoption of the 2004/2017 UNESCO definition of literacy, including both literacy and numeracy.
- Adoption of the PIAAC assessment framework as the global framework.
- Using the PIAAC assessment framework as the basis for further mapping of national assessment frameworks adding missing domains and sub-domains to fill the gaps in PIAAC framework. If possible, expanding the assessment framework of PIAAC to capture lower levels of the literacy and numeracy continuum.

The main outcomes of the meeting will be reported to the GAML Task Force 4.6 which will be convened on 28–29 November 2017 in Madrid, Spain. Meanwhile, mapping and further analysis of existing literacy assessments will be conducted. This expert group will be kept informed of developments on the indicator measurement and reporting, and will be consulted and convened electronically or in person on a need basis. Another meeting of this kind may be organized in 2018.

ANNEX

List of Participants

Silvia Montoya	Director	UNESCO Institute for Statistics (UIS)
Margarete Sachs-Israel	Chief Programme Coordinator	UNESCO Institute for Lifelong Learning (UIL)
William Thorn	Senior Analyst	OECD
Rakhat Zholdoshalieva	Programme Coordinator, Literacy and Basic Skills	UNESCO Institute for Lifelong Learning (UIL)
Anuja Singh	Programme Specialist	UNESCO Institute for Statistics (UIS)
Marta Encinas-Martin	Programme Manager	OECD
Jordan Naidoo	Director	Division for Education 2030 Support and Coordination, UNESCO
Manos Antoninis	Director	Global Education Monitoring Report, UNESCO
Astrid Gillet	Chief	Capacity Development and Field Support Unit, UNESCO
Koji Miyamoto	Senior Economist	World Bank Group
A.K.M Ashraful Haque	Project Director	Bangladesh Bureau of Statistics
Hakima Meghrebi	Professor	University of Paris XIII Nord, France
Jean-Pierre Jeantheau	Policy Officer	Nationale de Lutte contre l'Illettrisme (ANLCI), France
Anke Grotlüschen	Professor	University of Hamburg, Germany
Iddo Gal	Senior Lecturer	University of Haifa, Israel
Janet Chepkemoi Rotich	Deputy Director	The Directorate of Adult and Continuing Education in Ministry of Education, Kenya
Bryan Maddox	Senior Lecturer	University of East Anglia, UK
Scott Murray	President	DataAngel Inc., Canada